# This Page Is Inserted by IFW Operations and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

### IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

### (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

## (19) World Intellectual Property Organization International Bureau



### 

### (43) International Publication Date 14 December 2000 (14.12.2000)

### **PCT**

## (10) International Publication Number WO 00/75279 A2

(51) International Patent Classification7:

(21) International Application Number:

C12N

(74) Agents: CARROLL, Peter, G. et al.; Medlen & Carroll,

PCT/US00/15728

(22) International Filing Date: 6 June 2000 (06.06.2000)

. .

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 09/328,925 9 June 1999 (09.06.1999) US

(71) Applicant (for all designated States except US): RE-GENTS OF THE UNIVERSITY OF MICHIGAN [US/US]; 3003 South State Street, Wolverine Tower, Ann Arbor, MI 48109-1280 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KURACHI, Kotoku [JP/JP]; 2211 Avalon Place, Ann Arbor, MI 48104 (US). KURACHI, Sumiko [JP/JP]; 2211 Avalon Place, Ann Arbor, MI 48104 (US). LLP, Suite 2200, 220 Montgomery Street, San Francisco, CA 94104 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

#### Published:

 Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A2

(54) Title: NUCLEOTIDE SEQUENCES FOR GENE REGULATION AND METHODS OF USE THEREOF

(57) Abstract: The invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest. In particular, the invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner and/or in a liver-specific manner. The invention further provides methods of using the regulatory nucleic acid sequences provided herein for age-related and/or liver-specific expression of nucleotides sequences of interest. The invention also provides host cells and transgenic non-human animals which harbor the regulatory nucleic acid sequences of the invention. The compositions and methods of the invention are useful in regulating expression of a nucleotide sequence of interest in an age-related and/or liver-specific manner.

## NUCLEOTIDE SEQUENCES FOR GENE REGULATION AND METHODS OF USE THEREOF

This work was made, in part, with Government support by the National Institutes of Health grant numbers HL38644 and HL53713. The Government has certain rights in this invention.

### FIELD OF THE INVENTION

5

10

15

20

25

30

The invention relates to nucleic acid sequences which regulate expression of a nucleotide sequence of interest. In particular, the invention relates to nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner and/or in a liver-specific manner. The invention further relates to methods of using the regulatory nucleic acid sequences provided herein for age-related and/or liver-specific expression of nucleotides sequences of interest. The invention also relates to host cells and to transgenic non-human animals which harbor the regulatory nucleic acid sequences of the invention. The compositions and methods of the invention are useful in regulating expression of a nucleotide sequence of interest in an age-related and/or liver-specific manner.

### **BACKGROUND OF THE INVENTION**

A multitude of human diseases (e.g., thrombosis, cardiovascular diseases, diabetes, Alzheimer's disease, cancer, osteoporosis, osteoarthritis, Parkinson's disease, dementia) are associated with increasing age and result in serious effects on the quality of life and on the life expectancy of individuals suffering from such diseases. Other diseases (e.g., cirrhosis, primary and metastatic neoplasia, Wilson disease, hepachromatosis, infectious hepatitis, hepatic necrosis, Gilbert disease, Criggler-Najar disease) which afflict the liver also have serious clinical manifestations and are responsible for high morbidity and mortality.

The treatment of age-related diseases (i.e., diseases whose prevalence and/or severity of clinical manifestations increases with the age of the patient) and diseases afflicting the liver focuses on the alleviation of the general symptoms of the disease using one or a combination of two modalities, i.e., non-pharmacological treatment and pharmacological treatment. Non-pharmacological treatment include, for example, periods of bed rest and dietary changes. Non-pharmacological treatment is often used as an adjunct to

pharmacological treatment which involves the use of drugs. Unfortunately, many of the commonly used pharmacological agents have numerous side effects and their use is further exacerbated by the non-responsiveness by many patients with severe disease, who, paradoxically, are in most need of treatment. Both non-pharmacological and pharmacological treatments provide unsatisfactory approaches to treating age-related and liver-associated diseases because these approaches are often ineffective, their effects are inconsistent, and are directed to alleviating the general symptoms of disease, rather than to specifically addressing the source of morbidity and mortality. Moreover, no suitable animal models are currently available to rationally design drugs which target specific biochemical and physiological pathways which are associated with age-related and with liver-associated diseases.

What is needed are methods for age-related and liver-specific gene expression and models for age-related and liver-specific diseases.

### SUMMARY OF THE INVENTION

5

10

15

20

25

30

The invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner, as well as nucleic acid sequences which direct liver-specific expression of a gene of interest. Further provided by the invention are transgenic animals which may be used as models for age-related and/or liver specific diseases.

In one embodiment, the invention provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to any particular type or source of nucleic acids sequence of interest, in a preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase. While it is not intended that the invention be restricted to any particular type or source of promoter sequence, in an alternative preferred embodiment, the promoter sequence is selected from human factor IX

promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>7</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. It is not contemplated that the invention be limited to any particular age regulatory sequence which is a portion of SEQ ID NO:1. However, in another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:31 is selected from SEQ ID NO:2, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, and SEQ ID NO:38. Without intending to limit the invention to any particular rage regulatory sequence which s a portion of SEQ ID NO:3, in yet another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:3 is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

5

10 .

15

20

25

30

Also provided by the invention is a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to the environment in which the host cell is contained, in one preferred embodiment, the host cell is comprised in a tissue or organ in a living animal. In another preferred embodiment, the host cell is a gamete. In yet another preferred embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

The invention also provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without limiting the invention to the type or source of the nucleic acid sequence of interest, in one preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase. While it is not intended that the invention be limited to the type or source of the promoter sequence, in an alternative

preferred embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>7</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. Though it is not contemplated that the invention be limited to the portion of SEQ ID NO:1 which has age-related regulatory activity, in another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:1 is selected from SEQ ID NO:2, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, and SEQ ID NO:38. Without intending to limit the invention the portion of SEQ ID NO:3 which has age-related regulatory activity, in yet another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:3 is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58; SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

5

10

15

20

25

30

Also provided herein is a host cell containing recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to the environment in which the host cell is contained, in one preferred embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative preferred embodiment, the host cell is a gamete. In another preferred embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

The invention also provides a method, comprising: a) providing: i) a cell, ii) a nucleic acid sequence of interest, iii) a promoter sequence, and iv) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the one or more age regulatory sequences to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. Without intending to limit the treated cell to any particular environment, in one preferred embodiment, the treated cell is comprised in a tissue or organ in a living animal.

The invention further provides a substantially purified nucleic acid sequence comprising a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof.

Also provided herein is a substantially purified nucleic acid sequence comprising a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof.

5

10

15

20

25

30

Also provided by the present invention is a substantially purified nucleic acid sequence comprising a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof. In one embodiment, the portion is SEQ ID NO:2. In an alternative embodiment, the portion is selected from SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37 and SEQ ID NO:38.

The invention also provides a substantially purified nucleic acid sequence comprising a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof. In one embodiment, the portion is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with SEQ ID NO:1 or with the complement thereof, wherein the nucleic acid sequence is characterized by having age-related regulatory activity, and by having greater than 63% and less than 100% homology to the SEQ ID NO:1.

The invention also provides a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with SEQ ID NO:3 or with the complement thereof, wherein the nucleic acid sequence is characterized by having age-related regulatory activity, and by having greater than 60% and less than 100% homology to the SEQ ID NO:3.

The invention additionally provides a recombinant expression vector comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof.

Also provided herein is a recombinant expression vector comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof.

The invention also provides a transgenic cell comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof. In one embodiment, the nucleotide sequence is operably linked to a promoter and to a nucleic acid sequence of interest. In a preferred embodiment, the transgenic cell is comprised in an animal. In a more preferred embodiment, the nucleic acid sequence of interest is expressed in an age-related manner in the transgenic cell.

5

10

15

20

25

30

The invention additionally provides a transgenic cell comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof. In one embodiment, the nucleotide sequence is operably linked to a promoter and to a nucleic acid sequence of interest. In a preferred embodiment, the transgenic cell is comprised in an animal. In a more preferred embodiment, the nucleic acid sequence of interest is expressed in an age-related manner in the transgenic cell.

The invention also provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; iv) SEQ ID NO:1; and v) SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, the SEQ ID NO:1 and the SEO ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell. In one embodiment, the cell expresses a recombinant protein identified as SEQ ID NO:47. In an alternative embodiment, the cell is selected from HepG2 cell, fibroblast cell, myoblast cell, and endothelial cell. In another embodiment, the cell is a fertilized egg cell, and the transgenic cell is a transgenic fertilized egg cell. In a preferred embodiment, the method further comprises d) introducing the transgenic fertilized egg cell into a non-human animal and permitting the animal to deliver progeny containing the transgene. In a more preferred embodiment, the progeny is characterized by age-related expression of the nucleic acid sequence of interest. In an alternative more preferred embodiment, the progeny is characterized by liver-specific expression of the nucleic acid sequence of interest. In another preferred embodiment, the fertilized egg cell is derived from a mammal of the order Rodentia. In a more preferred embodiment, the fertilized egg cell is a mouse fertilized egg cell. In yet another embodiment, the promoter is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK

promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>7</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In a further embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase.

5

10

15

20

25

30

The invention also provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; iv) a portion of SEQ ID NO:1; and v) a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, the portion of SEQ ID NO:1 and the portion of SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Additionally provided by the invention is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) SEQ ID NO:1; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the SEQ ID NO:1 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Also provided herein is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a portion of SEQ ID NO:1; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the portion of SEQ ID NO:1 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

The invention further provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Further provided by the invention is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the portion of SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

5

10

15

20

25

30

The invention further also provides additional sequences derived from the hFIX gene. In particular, the invention provides a substantially purified nucleic acid sequence comprising at least a portion of SEQ ID NO:93. In one embodiment, the portion has age-related regulatory activity. In an alternative embodiment, the portion is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In a preferred embodiment, the portion is SEQ ID NO:91.

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under high stringency hybridization conditions with a nucleotide sequence selected from SEQ ID NO:91, the complement of SEQ ID NO:91, SEQ ID NO:93, and the complement of SEQ ID NO:93. In one embodiment, the nucleic acid sequence has agerelated regulatory activity.

The invention also provides a substantially purified nucleic acid sequence comprising a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof.

The invention further provides expression vectors containing hFIX sequences. In one embodiment, the invention provides a recombinant expression vector comprising in operable

5

10

15

20

25

30

combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) an agerelated regulatory sequence selected from SEQ ID NO:93 and portions thereof. In a preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In an alternative embodimnt, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRa promoter, P450 family promoters, GAL7 promoter, T<sub>2</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another alternative embodiment, the portion of SEO ID NO:93 is selected from SEO ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEO ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In a more preferred embodiment, the portion is SEQ ID NO:91. In another embodiment, the expression vector further comprises in operable combination an age-related regulatory sequence selected from SEQ ID NO:1 and portions thereof.

The invention further provides cells containing hFIX-derived sequences. In particular, the invention provides a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative embodiment, the host cell is a gamete. In another alternative embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

5

10

15

20

25

30

The invention additionally provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of an age-related regulatory sequence selected from SEO ID NO:93 and portions thereof. In one embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human \( \alpha 1-\) antitrypsin. antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In an alternative embodiment, the promoter sequence is selected from human factor IX promoter. cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRa promoter, P450 family promoters, GAL7 promoter, T<sub>2</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another embodiment, the portion of SEQ ID NO:93 is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEO ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In yet a further embodiment, the portion is SEQ ID NO:91. In another embodiment, the expression vector further comprises in operable combination an age-related regulatory sequence selected from SEQ ID NO:1 and portions thereof.

Also provided herein is host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In another embodiment, the host cell is a gamete. In yet

another embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

Also provided herein are methods for using hFIX-derived sequences. For example, the invention provides a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the age-related regulatory sequence to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. In one embodiment, the treated cell is comprised in a tissue or organ in a living animal.

5

10

15

20

25

30

The invention also provides a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the functional homolog to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell.

In addition to sequences from the hFIX gene, the invention provides sequences derived from the hPC gene. In particular, the invention provides substantially purified nucleic acid sequence comprising a nucleotide sequence selected from at least a portion of SEQ ID NO:85, and at least a portion of SEQ ID NO:92. In one embodiment, the portion has activity selected from age-related regulatory activity and regulatory activity. In an alternative embodiment, the portion is selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162; SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:179, SEQ ID NO:171, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:179, SEQ ID NO:181, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEO ID NO:184, SEO ID

NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In another alternative embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

5

10

15

20

25

30

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with a nucleotide sequence selected from SEQ ID NO:92, the complement of SEQ ID NO:92, SEQ ID NO:85, the complement of SEQ ID NO:85, SEQ ID NO:89, the complement of SEQ ID NO:89, SEQ ID NO:90, and the complement of SEQ ID NO:90. In one embodiment, the nucleic acid sequence has activity selected from age-related regulatory activity and regulatory activity.

The invention additionally provides a substantially purified nucleic acid sequence comprising a functional homolog of a sequence having activity selected from age-related regulatory activity and regulatory activity, the sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90.

Also provided by the invention is a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the expression vector of Claim 38, wherein the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In another embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>7</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another embodiment, the portion is

selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In a further embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

The invention also provides a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In another embodiment, the host cell is a gamete. In a further embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

25

30

5

10

15

20

The invention also provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin,

5

10

15

20

25

30

protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In another embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRa promoter, P450 family promoters, GAL7 promoter, T<sub>2</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In a further embodiment, the portion is selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In an alternative embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

Further provided herein is a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative embodiment, the host cell is a gamete. In yet another embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

Also provided herein are methods for using hPC sequences. In particular, the invention discloses a method for expressing a nucleic acid sequence of interest, comprising:

a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the nucleotide sequence to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. In one embodiment, the treated cell is comprised in a tissue or organ in a living animal.

Also provided herein is a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the functional homolog to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

5

10

15

20

25

30

Figure 1 shows the structure of eleven exemplary human FIX minigene expression constructs and relative *in vitro* transient expression activities (ng hFIX/10<sup>6</sup> cells/48 hr).

Figure 2 shows graphs of longitudinal analyses of transgenic mice which carry -416FIXm1 (A), -416FIXm1/1.4 (B), -590FIXm1 (C), -679FIXm1 (D), and -770FIXm1 (E) expression vectors and which produce high initial prepubertal, but rapidly decreasing, hFIX expression levels with age.

Figure 3 shows a Northern blot of human FIX mRNA levels (A) and a gel showing hFIX transgene DNA levels as determined by multiplex PCR analysis (B) in the livers and tails of animals carrying -416FIXm1.

Figure 4 shows graphs of longitudinal analysis of transgenic mice which carry -802FIXm1 (A), -802FIXm1/1.4 (B), -2231FIXm1 (C), -2231FIXm1/1.4 (D) and -416FIXm1/AE5' (E) expression vectors and which produce hFIX at stable and increasing levels with age.

Figure 5 shows a Northern blot of transgenic mice carrying -802FIXm1 and -802FIXm1/1.4 expression vectors.

5

10

15

20

25

30

Figure 6 is a gel of a gel electrophoretic mobility shift assay using mouse liver nuclear extract (NEs) from three different age groups, and using double-stranded oligonucleotides containing a PEA-3 element nucleotide sequence spanning from nt -797 to -776 of the hFIX gene (A), and using a competition assay for <sup>32</sup>P-labelled double stranded oligonucleotides containing the PEA-3 nucleotide sequence (B).

Figure 7 is a Northern blot showing tissue specificity of hFIX expression in transgenic mice carrying -416FIXm1 (A) and -802 FIXm1 (B) expression vectors.

Figure 8A-E shows the nucleotide sequence (SEQ ID NO:4) of, and eight amino acid sequences (SEQ ID NOs:5 to 12) which together form, the human factor IX (GenBank accession number K02402). The initiation transcription site (nucleotide 1) and the poly-A addition site (nucleotide 32,757) are identified by solid circles. The solid vertical arrows indicate the intron-exon splice junction. The five Alu repetitive sequences have been underlined, while the 5-base insert in intron A and the AATAAA sequence in exon VIII are boxed. The cleavage or termination site at the 3' end of the gene (CATTG) is underlined with a dashed line.

Figure 9 shows the cDNA sequence (SEQ ID NO:13) (A) and encoded polypeptide sequence (SEQ ID NO:47) (B) of mouse PEA-3 (GenBank accession number X63190).

Figure 10 A-D shows the cDNA sequence (SEQ ID NO:42) of the human  $\alpha$ 1-antitrypsin gene (GenBank accession number K02212).

Figure 11 shows the DNA sequence (SEQ ID NO:43) of human antithrombin III (GenBank accession number A06100).

Figure 12 shows the cDNA sequence (A) (SEQ ID NO:49) (GenBank accession number X02750) and genomic DNA sequence (B) (SEQ ID NO:50) (GenBank accession number M11228) of human protein C.

Figure 13 (A-E) shows the nucleic acid sequences (SEQ ID NOs:76-83) of exemplary homologs of AE3' (SEQ ID NO:3).

Figure 14 shows the nucleotide sequence (from nt -1462 to nt +1; SEQ ID NO:85) which is located at the 5'-end of the human protein C gene.

Figure 15 shows the structure of eight exemplary human protein C minigene expression constructs.

Figure 16 shows the relative *in vitro* transient expression activities for five exemplary human protein C minigene expression constructs.

Figure 17 shows graphs of longitudinal analyses of transgenic mice which carry -1462hPCm1 (A), -82hPCm1 (B), and AE5'/-1462hPCm1/AE3' (C) expression vectors.

### **DEFINITIONS**

5

10

15

20

25

30

To facilitate understanding of the invention, a number of terms are defined below.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated nucleic acid sequence" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural state, or when obtained from its actual source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, an isolated nucleic acid sequence comprising SEQ ID NO:1 includes, by way of example, such nucleic acid sequences in cells which ordinarily contain SEQ ID NO:1 where the nucleic acid sequence is in a chromosomal or extrachromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid sequence may be present in single-stranded or double-stranded form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain (at a minimum) at least a portion of the sense or coding strand (i.e., the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and antisense strands (i.e., the nucleic acid sequence may be double-stranded).

As used herein, the term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An

"isolated nucleic acid sequence" is therefore a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated.

The term "recombinant" when made in reference to a DNA sequence refers to a DNA sequence which is comprised of segments of DNA joined together by means of molecular biological techniques. The term "recombinant" when made in reference to a polypeptide sequence refers to a polypeptide sequence which is expressed using a recombinant DNA sequence.

5

10

15

20

25

30

As used herein, the terms "vector" and "vehicle" are used interchangeably in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another.

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes include a promoter, optionally an operator sequence, a ribosome binding site and possibly other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

The term "transgenic" when used in reference to a cell refers to a cell which contains a transgene, or whose genome has been altered by the introduction of a transgene. The term "transgenic" when used in reference to a tissue or animal refers to a tissue or animal, respectively, which comprises one or more cells that contain a transgene, or whose genome has been altered by the introduction of a transgene. Transgenic cells, tissues and animals may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into a target cell or integration of the transgene into a chromosome of a target cell by way of human intervention, such as by the methods described herein.

A "non-human animal" refers to any animal which is not a human and includes vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc. Preferred non-human animals are selected from the order Rodentia. The term "order Rodentia" refers to rodents *i.e.*, placental mammals (class Euthria) which include the family Muridae (e.g., rats and mice), most preferably mice.

The term "nucleotide sequence of interest" refers to any nucleotide sequence, the manipulation of which may be deemed desirable for any reason (e.g., treat disease, confer improved qualities, etc.), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (e.g., reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factors, etc.), and noncoding regulatory sequences which do not encode an mRNA or protein product (e.g., promoter sequence, polyadenylation sequence, termination sequence, enhancer sequence, etc.).

5

10

15

20 -

25

30

As used herein, the terms "complementarity," or "complementary" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

A "complement" of a nucleic acid sequence as used herein refers to a nucleotide sequence whose nucleic acids show total complementarity to the nucleic acids of the nucleic acid sequence.

The term "homology" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology (i.e., partial identity) or complete homology (i.e., complete identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid sequence and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (i.e., an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence to a target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific

binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

5

10

15

20

25

30

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize to the single-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

The term "hybridization" as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing." [Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY]. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T<sub>m</sub> of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein, the term " $T_m$ " is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the  $T_m$  of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the  $T_m$  value may be calculated by the equation:  $T_m = 81.5 + 0.41(\% G + C)$ , when a nucleic acid is in aqueous solution at 1 M NaCl [see e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985)]. Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of  $T_m$ .

Low stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH<sub>2</sub>PO<sub>4</sub>•H<sub>2</sub>O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 1% SDS, 5X Denhardt's reagent [50X Denhardt's contains the following per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml

denatured salmon sperm DNA followed by washing in a solution comprising 0.2X SSPE, and 0.1% SDS at room temperature when a DNA probe of about 100 to about 1000 nucleotides in length is employed.

5

10

15

20

25

30

High stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE, 1% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, and 0.1% SDS at 68°C when a probe of about 100 to about 1000 nucleotides in length is employed.

The term "equivalent" when made in reference to a hybridization condition as it relates to a hybridization condition of interest means that the hybridization condition and the hybridization condition of interest result in hybridization of nucleic acid sequences which have the same range of percent (%) homology. For example, if a hybridization condition of interest results in hybridization of a first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence, then another hybridization condition is said to be equivalent to the hybridization condition of interest if this other hybridization condition also results in hybridization of the first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence.

When used in reference to nucleic acid hybridization the art knows well that numerous equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency hybridization different from, but equivalent to, the above-listed conditions.

Those skilled in the art know that whereas higher stringencies may be preferred to reduce or eliminate non-specific binding of the nucleotide sequence of SEQ ID NOs:1 or 3 with other nucleic acid sequences, lower stringencies may be preferred to detect a larger number of nucleic acid sequences having different homologies to the nucleotide sequence of SEQ ID NOs:1 and 3.

As used herein, the terms "regulatory element" and "regulatory sequence" interchangeably refer to a nucleotide sequence which does not encode RNA or a protein and which controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, etc. In contrast, the term "regulatory gene" refers to a DNA sequence which encodes RNA or a protein (e.g., transcription factor) that controls the expression of other genes.

5

10

15

20

25

30

Regulatory elements may be tissue specific or cell specific. The term "tissue specific" as it applies to a regulatory element refers to a regulatory element that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (e.g., liver) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (e.g., lung).

Tissue specificity of a regulatory element may be evaluated by, for example, operably linking a reporter gene to a promoter sequence (which is not tissue-specific) and to the regulatory element to generate a reporter construct, introducing the reporter construct into the genome of an animal such that the reporter construct is integrated into every tissue of the resulting transgenic animal, and detecting the expression of the reporter gene (e.g., detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic animal. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the regulatory element is "specific" for the tissues in which greater levels of expression are detected. Thus, the term "tissue-specific" (e.g., liver-specific) as used herein is a relative term that does not require absolute specificity of expression. In other words, the term "tissue-specific" does not require that one tissue have extremely high levels of expression and another tissue have no expression. It is sufficient that expression is greater in one tissue than another. By contrast, "strict" or "absolute" tissue-specific expression is meant to indicate expression in a single tissue type (e.g., liver) with no detectable expression in other tissues.

The term "cell type specific" as applied to a regulatory element refers to a regulatory element which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide

sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a regulatory element also means a regulatory element capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue.

5

10

15

20

25

30

Cell type specificity of a regulatory element may be assessed using methods well known in the art, e.g., immunohistochemical staining and/or Northern blot analysis. Briefly, for immunohistochemical staining, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is regulated by the regulatory element. A labeled (e.g., peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (e.g., with avidin/biotin) by microscopy. Briefly, for Northern blot analysis, RNA is isolated from cells and electrophoresed on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled oligo-deoxyribonucleotide probe or DNA probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists.

The term "promoter," "promoter element," or "promoter sequence" as used herein, refers to a DNA sequence which when ligated to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. A promoter is typically, though not necessarily, located 5' (i.e., upstream) of a nucleotide sequence of interest whose transcription into mRNA it controls, and provides a site for specific binding by RNA polymerase and other transcription factors for initiation of transcription.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (e.g., heat shock, chemicals, etc.). In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a stimulus (e.g., heat shock, chemicals, etc.) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The terms "essentially consisting of" and "consisting essentially of" are equivalent terms, and when in reference to a nucleic acid sequence they are intended to refer to nucleotide sequences which contain from 50% to 100% of the nucleic acid bases which are present in the nucleic acid sequence, in which the arrangement of these nucleic acid bases with respect to each other in the nucleotide sequences is the same as their arrangement in the nucleic acid sequence, and in which the biological activity of the nucleotide sequences is from 50% to 100%, more preferably from 75% to 100%, and most preferably from 90% to 100%, of the biological activity of the nucleic acid sequence. To illustrate, the term "a nucleic acid sequence consisting essentially of SEQ ID NO:1" refers to nucleotide sequences which contain from 50% to 100% of the nucleic acid bases which are present in SEQ ID NO:1, in which the arrangement of these nucleic acid bases with respect to each other in the nucleotide sequences is the same as their arrangement in SEQ ID NO:1, and in which the nucleotide sequences exhibit from 50% to 100%, more preferably from 75% to 100%, and most preferably from 90% to 100%, of the age-related regulatory activity, and/or of the liver-specific activity of SEQ ID NO:1.

5

10

15

20

25

30

A "functional homolog" of a nucleotide sequence which is derived from the hFIX gene shown in Figure 8 is defined as a nucleic acid sequence which has more than 50% identity and less than 100% identity with the hFIX-derived nucleotide sequence (i.e., with the entire, or a portion of the, hFIX sequence of Figure 8), and which has age-related regulatory activity and/or liver-specific activity. For example, a functional homolog of SEQ ID NO:33 includes nucleic acid sequences which have more than 50% identity and less than 100% identity with SEQ ID NO:33, and which have age-related regulatory activity and/or liver-specific activity.

A "functional homolog" of a nucleotide sequence which is derived from the hPC sequence shown in Figure 14 is defined as a nucleic acid sequence which has more than 50% identity and less than 100% identity with the hPC-derived nucleotide sequence (i.e., with the entire, or a portion of the, hPC sequence of Figure 14), and which has age-related regulatory activity and/or regulatory activity. For example, a functional homolog of SEQ ID NO:93 includes nucleic acid sequences which have more than 50% identity and less than 100% identity with SEQ ID NO:93, and which have age-related regulatory activity and/or regulatory activity.

### **DESCRIPTION OF THE INVENTION**

5

10

15

20

25

The invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest. In on embodiment, the invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner. Yet more particularly, the exemplary age-regulatory element 5' (AE5') has been discovered to regulate stable gene expression over time in vivo, while the exemplary ageregulatory element 3' (AE3') has been discovered to regulate increased gene expression over time in vivo. In another embodiment, the invention provides nucleic acid sequences (e.g., AE5') which direct liver-specific expression of a gene of interest. In yet another embodiment, the invention provides transgenic animals which harbor the nucleic acid sequences provided herein and which expres a nucleotide sequence of interest in an agerelated and/or liver-specific manner. The nucleic acid sequences provided herein are useful in, for example, identifying and isolating functional homologs of AE5' and AE3', and amplifying at least a portion of AE5' and AE3'. Importantly, the nucleic acid sequences of the invention are also useful in age-related expression and/or liver-specific expression of a nucleotide sequence of interest in an animal, in gene therapy, and in reducing expression of factor IX in an animal.

In addition to regulatory sequences which are derived from the human factor IX gene (e.g., AE5', AE3', and AE3''), the invention further provides sequences which are derived from the human protein C (hPC) gene and which are characterized by age-related regulatory activity and regulatory activity.

The invention is further discussed under (A) Regulatory Nucleic Acid Sequences, (B) Using Probes To Identify And Isolate Homologs Of AE5', AE3', and Of hPC-Derived Regulatory Sequences, (C) Using Primers to Amplify At Least A Portion Of AE5', AE3', and Of hPC-Derived Regulatory Sequences, (D) Methods For Regulating Gene Expression, (E) Gene Therapy, and (F) Reducing Expression Of Factor IX In An Animal.

### A. Regulatory Nucleic Acid Sequences

5

10

15

20

25

30

The invention provides regulatory sequences which are derived from the hFIX and hPC genes.

### i. Regulatory Nucleic Acid Sequences From The hFIX Gene

The regulatory nucleic acid sequences of the invention and their surprising properties in regulating gene expression were discovered during the inventor's investigation of the mechanisms underlying age-associated regulation of the human factor IX, which is involved in blood coagulation. Blood coagulation plays a critical role not only in homeostasis, but also in many physiological and pathological conditions [Saito in Disorders of Hemostasis, O.D. Ratnoff and C.D. Forbes, Eds., Sauders, Philadelphia, ed. 2 (1991), pp. 18-47; Kurachi et al. (1993) Blood Coagul. Fibrinol. 4:953-974]. Blood coagulation potential in humans as well as in other mammals reaches the young adult level around the age of weaning [Yao et al. (1991) Thromb. Haemost. 65:52-58; Andrew et al. (1992) Blood 80:1998-2005; Andrew et al. Blood (1987) 70:165-172; Andrew et al. (1988) Blood 72:1651-1657]. This is followed by a gradual increase in coagulation potential during young adulthood, and an almost two-fold increase by old age [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi et al. (1996) Thromb. Haemost. 76:965-969]. This age-associated increase in coagulation potential takes place in healthy centenarians [Marie et al. (1995) Blood 85:3144-3149], indicating that the increase is a normal phenomenon associated with aging.

It is the inventors' consideration that this increase in coagulation potential may make a crucial contribution to the development and progression of age-associated diseases such as cardiovascular and thrombotic disorders [Conlan et al. (1993) The Atherosclerosis Risk in Communities (ARIC) Study 70:380-385; Balleisen et al. (1985) Thromb. Haemost. 54:475-479; Rode et al. (1996) Nat. Med. 2:293-298; Woodward et al. (1997) Brit. J. Haemat. 97:785-797]. The inventors' consideration was based on the observation that this increase in blood coagulation potential coincides with plasma level increases of pro-coagulant factors such as factor IX, whereas plasma levels of anti-coagulation factors (such as antithrombin III and protein C) or of factors involved in fibrinolysis are only marginally affected [Conlan et al. (1994) The Atherosclerosis Risk in Committees (ARIC) Study 72:551-556; Lowe et al. (1997) Brit. J. Haemat. 97:775-784]. These facts strongly suggested to the inventors that the

observed increase in blood coagulation activity with advancing age is due to regulated events. Plasma levels of each protein factor involved in blood coagulation, fibrinolysis and their regulatory systems are presumably determined by the balance of the many processes involved. At present, little is known about why an advancing age-associated increase in blood coagulation activity exists, or what molecular mechanisms are involved in age-dependent regulation (homeostasis) of blood coagulation [Finch in *Longevity, Senescence, and the Genome*, The University of Chicago Press, Chicago, 1990].

5

10

15

20

25

30

Blood coagulation factor IX (FIX) occupies a key position in the blood coagulation cascade where the intrinsic and extrinsic pathways merge [Saito in *Disorders of Hemostasis*, O.D. Ratnoff and C.D. Forbes, Eds., Sauders, Philadelphia, ed. 2 (1991), pp. 18-47; Kurachi et al. (1993) Blood Coagul. Fibrinol. 4:953-974]. FIX is synthesized in the liver with strict tissue-specificity, and its deficiency results in the bleeding disorder hemophilia B. In normal humans the plasma activity and protein concentration levels of human FIX (hFIX) increase with advancing age [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi et al. (1996) Thromb. Haemost. 76:965-969]. Mouse FIX (mFIX) plasma activity also increases with age in a manner similar to hFIX, and is directly correlated with an increase in liver mFIX messenger RNA (mRNA) level [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi et al. (1996) Thromb. Haemost. 76:965-969]. However, nothing else is known about the molecular mechanisms underlying such an increase. In investigating the basic molecular mechanisms responsible for age-associated regulation of hFIX, the inventors discovered the nucleotide sequences which regulate age-associated expression, and which direct liver-specific expression, of the exemplary hFIX gene.

The discovery of the invention sequences was made possible, in part, by the inventors' use of the hFIX promoter in combination with the coding sequence for hFIX instead of with the coding sequence for commonly used reporter proteins. The discovery of the surprising functions of the nucleotide sequences provided herein was also made possible by the inventors' use of longitudinal *in vivo* analyses, rather than of *in vitro* analyses. In particular, the inventors' earlier studies used reporter genes (including bacterial β-galactosidase and chloramphenicol acetyltransferase [CAT]) which are heterologous to the factor IX promoter. In these earlier studies, the factor IX promoter showed only very weak expression activity *in vitro* [Kurachi et al. (1995) J. Biol. Chem. 270:5276-5281]. Use of such heterologous reporter genes made it impossible to reliably and quantitatively perform

longitudinal analyses of transgene expression in animals. The inventors unexpectedly observed that the use of hFIX minigene expression vectors which contained the hFIX promoter and its homologous hFIX gene were capable of producing high level plasma hFIX in vivo. This unexpected observation not only solved the problems associated with the use of genes which are heterologous to the hFIX promoter by providing a reliable animal assay system, but also provided multiple unexpected critical insights into the regulatory mechanisms of the hFIX gene, including the determination of nucleotide sequences which regulate the stability and age-related increased expression of the exemplary hFIX gene.

5

10

15

20

25

30

The present invention provides the 32-nucleotide nucleic acid sequence 5'-agccatt cagtcgagga aggatagggt ggtat-3' (SEQ ID NO:1) of AE5' which corresponds to the sequence from 2164 to 2195 of the hFIX gene deposited in GenBank as accession number K02402, and which corresponds to the sequence from -802 to -771 of Figure 8 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30.

The present invention also provides the 1273-nucleotide nucleic acid sequence (SEQ ID NO:3) (Figure 13) of AE3' which corresponds to the sequence from 34,383 to 35,655 of GenBank accession number K02402, and which corresponds to the sequence from 31,418 to 32,690 of Figure 8 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30.

The terms "age-related regulatory activity" and "age-related activity" when made in reference to a nucleic acid sequence refer to the ability of the nucleic acid sequence to alter in an age-related manner (e.g., increase over a period of time) the level of transcription into mRNA and/or the synthesis of a polypeptide encoded by a nucleotide sequence of interest which is operably linked to a promoter sequence as compared to the level of transcription into mRNA of the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has age-related regulatory activity. An "age regulatory sequence" is herein used to refer to a nucleic acid sequence which has age-related regulatory activity.

To illustrate, where expression levels of a gene of interest decrease over a period of time, a nucleic acid sequence is said to have age-related regulatory activity if (when operably linked to the gene of interest) it results in (a) a smaller decrease in expression levels of the gene over the same period of time as compared to the decrease in expression levels in the absence of the nucleic acid sequence, (b) relatively constant (i.e., unchanged) expression

levels over the same period of time, or (c) increased expression levels over the same period of time.

5

10

15

20

25

30

The terms "operably linked," "in operable combination," and "in operable order" as used herein refer to the linkage of nucleic acid sequences such that they perform their intended function. For example, operably linking a promoter sequence to a nucleotide sequence of interest refers to linking the promoter sequence and the nucleotide sequence of interest in a manner such that the promoter sequence is capable of directing the transcription of the nucleotide sequence of interest and/or the synthesis of a polypeptide encoded by the nucleotide sequence of interest. Similarly, operably linking a nucleic acid sequence having age-related regulatory activity to a promoter sequence and to a nucleotide sequence of interest means linking the nucleic acid sequence having age-related regulatory activity, the promoter sequence and the nucleotide sequence of interest in a manner such that the nucleic acid sequence having age-related regulatory activity is capable of altering over a period of time the level of transcription into mRNA of the nucleotide sequence of interest and/or the synthesis of a polypeptide encoded by the nucleotide sequence of interest.

Methods for determining age-related regulatory activity of a candidate nucleic acid sequence, given the teachings of the present specification, are within the ordinary skill in the art and are exemplified by the methods disclosed herein. For example, a test vector is constructed in which the candidate nucleic acid sequence is linked upstream or downstream of a promoter sequence which is operably linked to a nucleotide sequence of interest (e.g., Example 1). A control vector which is similar to the test vector but which lacks the candidate nucleic acid sequence is also constructed. The test vector and control vector are separately introduced into a host cell. It is preferred that the host cell (e.g., fertilized egg) be capable of generating a transgenic multicellular organism, e.g., a transgenic mouse (e.g., Example 3) and that transgenic multicellular organisms are generated. Longitudinal analyses of the expression of mRNA which is encoded by the nucleotide sequence of interest (e.g., by Northern blot hybridization) over a period of time in, and preferably over the entire life span of, founders and successive generations of the transgenic multicellular organism are carried out (e.g., Example 3). The detection in any tissue of mRNA and/or protein levels which are encoded by the nucleotide sequence of interest and which are greater in transgenic animals harboring the test vector as compared to the mRNA and/or protein levels in transgenic

animals harboring the control vector at least one point in time indicates that the candidate nucleic acid sequence has age-related regulatory activity.

For example, evidence provided herein shows the surprising result that AE5' (SEQ ID NO:1) alone has age-related regulatory activity in that AE5' stabilizes hFIX mRNA whereby hFIX mRNA levels are essentially unchanged at different time points over the entire life span of transgenic animals (Figure 4, A, C and E) as compared to the declining hFIX mRNA levels in transgenic animals which harbor vectors that lack AE5' (Figures 2A and 2E). The age-related regulatory activity of AE5' was observed regardless whether AE5' was placed upstream (Figure 4A) or downstream (Figure 4E) of the promoter sequence in the expression construct.

5

10

15

20

25

30

Furthermore, data provided herein demonstrates the unexpected result that AE3' (SEO ID NO:3) alone has age-related regulatory activity in that AE3' increases hFIX mRNA at several time points during the life of transgenic animals (Figure 2B) relative to the hFIX mRNA levels at the same time points in transgenic animals harboring vectors that lack AE3' (Figure 2A). AE3' substantially increased the steady state hFIX mRNA levels (Figure 5). This result which was observed in vivo was surprising in part because AE3' exhibited weak down regulatory effects on hFIX production in vitro. Without limiting the invention to any particular mechanism, these results suggest that AE3' functions by increasing hFIX mRNA stability which directly correlates with an increase in the hFIX protein level in the circulation. Also without intending to limit the invention to any particular theory, it is the inventors' consideration that the age-related regulatory activity of AE3' is due to the s1 structure-forming dinucleotide repeats present in the 3'UTR; the s1 region is the 103 bp sequence (SEQ ID NO:61) from nt 32,141 through nt 32,243 of Figure 8. This consideration is based on the inventors' observation that dinucleotide repeats, such as (AT), of the 3' UTR of various genes, can form sl structures in mRNA, which upon binding specific proteins are known to modulate mRNA stability, mostly to a less stable state [Ross (1995) Microbiol. Rev. 59:423-450].

Importantly, the invention demonstrates the surprising synergistic action of AE5' and AE3' which together result in hFIX mRNA levels which not only are greater at each time point tested over the life span of transgenic animals (Figures 4 B and D) as compared to hFIX mRNA levels in transgenic animals harboring vectors that lack both AE5' and AE3', but also that the profile of increased human FIX mRNA levels over the life span of

transgenic mouse recapitulates the profile of increased mouse FIX mRNA levels as a wild-type mouse ages.

Data presented herein further demonstrate that the age-related regulatory activity of AE5' alone, of AE3' alone, and of the combination of AE5' and AE3' is independent of the level of expression of the transgenes harboring them, sex, generation or zygosity status of the transgenic animals.

5

. 10

15

20

25

30

The present invention is not limited to SEQ ID NOs:1 and 3 but specifically contemplates portions thereof. As used herein the term "portion" when made in reference to a nucleic acid sequence refers to a fragment of that sequence. The fragment may range in size from five (5) contiguous nucleotide residues to the entire nucleic acid sequence minus one nucleic acid residue. Thus, a nucleic acid sequence comprising "at least a portion of" a nucleotide sequence comprises from five (5) contiguous nucleotide residues of the nucleotide sequence to the entire nucleotide sequence.

In a preferred embodiment, portions of SEQ ID NO:1 contemplated to be within the scope of the invention include, but are not limited to, the 7-nucleotide nucleic acid sequence of the polyomavirus enhance activator 3 (PEA-3) (5'-GAGGAAG-3') (SEQ ID NO:2) which corresponds to the sequence from 2176 to 2182 of GenBank accession number K02402, and which corresponds to the sequence from -790 to -784 of GenBank accession number K02402 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30. A nucleotide sequence [5'-CAGGAAG-3' (SEQ ID NO:40)] which is homologous to the invention's PEA-3 nucleotide sequence was initially reported in the art as a polyoma virus enhancer, and was reported to be involved in the regulation of expression of various genes (e.g., collagen gene and c-fos) in several tissues [Martin et al. (1988) Proc. Natl. Acad. Sci. 85:5839-5843; Xin et al. (1992) Genes & Develop. 6:481-496; Chotteau-Lelievre et al. (1997) Oncogene 15:937-952; Gutman and Wasylyk (1990) EMBO J. 9:2241-2246]. However, the PEA-3 protein sequence [or PEA-3 related protein(s)] which binds to nucleotide sequences which are homologous to the invention's PEA-3 nucleotide sequence has not been reported to be either liver-specific or enriched in the liver.

Other portions of SEQ ID NO:1 included within the scope of the invention include, for example, SEQ ID NO:33 [5'-tcgaggaagga-3'], SEQ ID NO:34 [5'-agtcgaggaaggata-3'], SEQ ID NO:35 [5'-tcagtcgaggaaggatagg-3'], SEQ ID NO:36 [5'-attcagtcgaggaaggatagggt-3'], SEQ ID NO:37 [5'-ccattcagtcgaggaaggatagggtgg-3'], and

SEQ ID NO:38 [5'-gccattcagtcgaggaaggatagggtggta-3'], all of which include the PEA-3 nucleotide sequence.

5

10

15

20

25

30

In a preferred embodiment, portions of SEO ID NO:3 contemplated to be within the scope of the invention include, but are not limited to, SEQ ID NO:51 [5'-TTATTTTATATATATATATATATATAAAATA-3'], SEQ ID NO:52 [5'-TAT AATATA-3'], SEQ ID NO:53 [5'-CAATATAAATATATAG-3'], SEQ ID NO:54 [5'-combination of SEQ ID NOs:51 and 52, i.e., SEQ ID NO:55 [5'-TTATTTTATA TATATATATATATAAAATATATAATATA-3'], the combination of SEQ ID NOs:52 and 53, i.e., SEQ ID NO:56 [5'-TATAATATACAATATAAATATAG-3'], the combination of SEQ ID NOs:53 and 54, i.e., SEQ ID NO:57 [5'-CAATATAAAT combination of SEQ ID NOs:51, 52, 53, and 54, i.e., SEQ ID NO:58 [5'-TTAT TTTATATATATATATATATAAAATATATAATATACAATATAAATATAGTGT GTGTGTATGCGTGTGTAGACACACACACACACACACATA-3'], the 723 bp sequence (SEQ ID NO:59) from nt 31,418 through nt 32,140 of Figure 8, the 447 bp sequence (SEO ID NO:60) from nt 32,244 through nt 32,690 of Figure 8, and the 103 bp sequence (SEQ ID NO:61) (i.e., the s1 region of the 3' UTR) from nt 32,141 through nt 32,243 of Figure 8.

The nucleotide sequence of portions of SEQ ID NOs:1 and 3 which exhibit agerelated regulatory activity may be determined using methods known in the art, e.g., using deletion constructs (e.g., see Yang et al. (1998) J. Biol. Chem. 273:891-897). Briefly, several expression plasmids are constructed to contain a reporter gene under the control of a promoter and of different candidate nucleotide sequences which are obtained either by restriction enzyme deletion of internal sequences in SEQ ID NOs:1 and 3, restriction enzyme truncation of sequences at the 5' and/or 3' end of SEQ ID NOs:1 and 3, by the introduction of single nucleic acid base changes by PCR into SEQ ID NOs:1 and 3, or by chemical synthesis. The gene-related regulatory activity of the different constructs is determined as described supra in order to determine whether the candidate nucleotide sequence exhibits age-related regulatory activity.

The sequences of the present invention are not limited to SEQ ID NOs:1 and 3 and portions thereof, but also include homologs of SEQ ID NOs:1 and 3, and homologs of portions thereof. Homologs of SEQ ID NOs:1 and 3, and of portions thereof, include, but

are not limited to, nucleotide sequences having deletions, insertions or substitutions of different nucleotides or nucleotide analogs as compared to SEQ ID NOs:1 and 3, and of portions thereof, respectively. Such homologs may be produced using methods well known in the art.

5

10

15

20

25

30

A "homolog" of SEQ ID NO:1 is defined as a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1. Homologs of SEQ ID NO:1 are exemplified, but not limited to, SEQ ID NO:66 (5'-acccatt cagtcgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the G at nt 2,165 is replaced with a C; SEQ ID NO:67 (5'-agccatt gagtcgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the C at nt 2,171 is replaced with a G; SEO ID NO:68 (5'agccatt cagacgagga aggatagggt ggtat-3') is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the T at nt 2,174 is replaced with a A; SEQ ID NO:69 (5'-agccatt cagtcgagga aggatagggt ggttt-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the A at nt 2,194 is replaced with a T; SEQ ID NO:70 (5'-agccatt cagtcgagga tcccaagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that AGGGT beginning at nt 2,186 is replaced with TCCCA; SEQ ID NO:71 (5'-agccatt cagtcgagga aggatagggcctaat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that TGGT beginning at nt 2,190 is replaced with CCTA; SEO ID NO:72 (5'-agaccatt cagtegagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that a A is inserted after nt 2,165; SEQ ID NO:73 (5'-agccatt cagtcgagga aggatagcggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that a C is inserted after nt 2,187; SEQ ID NO:74 (5'-agccatt cagtcgagga aggataat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that GGGTGGT beginning at nt 12,187 is deleted; and SEQ ID NO:75 (5'-agccatt cgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that CAGT beginning at nt 2,171 is deleted.

A "homolog" of SEQ ID NO:2 is defined as a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2. Homologs of SEQ ID NO:2 include, for example, GAGGATG (SEQ ID NO:39), CAGGAAG (SEQ ID NO:40),

CAGGATG (SEQ ID NO:41), GTGGAAG (SEQ ID NO:62), GTGGATG (SEQ ID NO:63), CTGGAAG (SEQ ID NO:64), CTGGATG (SEQ ID NO:65), and CAGGAAG (SEQ ID NO:84).

5

10

15

20

25

30

A "homolog" of SEO ID NO:3 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:3. Homologs of SEQ ID NO:3 are exemplified, but not limited to, SEQ ID NOs:76-83 shown in Figure 13. Specifically, SEQ ID NO:76 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the C at nt 34,390 has been replaced with a G. SEQ ID NO:77 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the T at nt 34,649 has been replaced with a A. SEQ ID NO:78 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the GC beginning at nt 34,959 has been replaced with a CG. SEQ ID NO:79 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the CATG beginning at nt 35,501 has been replaced with a GTAC. SEQ ID NO:80 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that TT is inserted after the A at nt 34,681. SEQ ID NO:81 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that TGC is inserted after the C at nt 35,581. SEQ ID NO:82 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that A at nt 35,636 is deleted. SEQ ID NO:83 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the G at nt 34,383 is deleted.

A "homolog" of SEQ ID NO:59 is defined as a nucleotide sequence having less than 100% and more than 62% identity with SEQ ID NO:59.

A "homolog" of SEQ ID NO:60 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:60.

A "homolog" of SEQ ID NO:61 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:61.

Homologs of a portion of SEQ ID NO:1 are exemplified by homologs of the PEA-3 nucleotide sequence (SEQ ID NO:2), which include, for example, GAGGATG (SEQ ID NO:39), CAGGAAG (SEQ ID NO:40), CAGGATG (SEQ ID NO:41), GTGGAAG (SEQ ID NO:62), GTGGATG (SEQ ID NO:63), CTGGAAG (SEQ ID NO:64), CTGGATG (SEQ ID NO:65), and CAGGAAG (SEQ ID NO:84).

The present invention also contemplates functioning or functional homologs of SEQ ID NO:1, of portions of SEQ ID NO:1 (e.g., functional portions of SEQ ID NOs:2, and 33-38), of SEQ ID NO:3, and of portions of SEQ ID NO:3 (e.g., functional portions of SEQ ID NOs:51-61).

A "functional homolog" of SEQ ID NO:1 is defined as a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1, and which has agerelated regulatory activity. Alternatively, a functional homolog of SEQ ID NO:1 is a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1, and having liver-specific activity.

5

10

15

20

25

30

A "functional homolog" of SEQ ID NO:2 is defined as a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2, and which has agerelated regulatory activity. Alternatively, a functional homolog of SEQ ID NO:2 is a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2, and having liver-specific activity.

A "functional homolog" of SEQ ID NO:3 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:3, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:59 is defined as a nucleotide sequence having less than 100% and more than 62% identity with SEQ ID NO:59, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:60 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:60, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:61 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:61, and which has age-related regulatory activity.

The present invention is not limited to sense molecules of SEQ ID NOs:1 and 3 but contemplates within its scope antisense molecules comprising a nucleic acid sequence complementary to at least a portion (e.g., a portion greater than 10 nucleotide bases in length and more preferably greater than 100 nucleotide bases in length) of the nucleotide sequence of SEQ ID NOs:1 and 3. These antisense molecules find use in, for example, reducing or

preventing expression of a gene (e.g. hFIX) whose expression is regulated by SEQ ID NOs:1 and 3.

The invention further provides the nucleotide sequence AE3" which is a preferred portion of AE3". AE3" [5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatata aatatatata taaaatatat aatatatata atgggagcaa taagcaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3"; (SEQ ID NO:93)] is the 154-nucleotide nucleic acid sequence from 35,075 to 35,228 of GenBank accession number K02402, which corresponds to the sequence from 32,110 to 32,263 of Figure 8 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30. AE3" contains a 102-bp stem-loop forming sequence (SEQ ID NO:91).

5

10

15

20

25

30

Data presented herein demonstrates the universality of the regulatory function of portions of AE3' in that AE3" has been successfully used to regulate expression of the heterologous hPC gene in an age-related manner. In particular, transgenic animals containing the -1462hPCm1 construct expressed age-stable levels of human protein C, *i.e.*, expressed relatively constant levels of human protein C at different time points during the life span of the transgenic animals (Example 12, Figure 17A). In direct contrast, the presence of AE5' and AE3'' sequences resulted in increased expression levels of human protein C over time (Figure 17C). These results confirm the universality of the function of the AE3'' portion of AE3' in regulating expression of operably linked genes in an age-related manner.

The invention also contemplates portions of AE3". These portions include, but are not limited to, SEQ ID NOs:94-144 wherein SEQ ID NO:94 is 5'-tgggg gaaaagtttc tttcagagag ttaagttatt ttatatata aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:95 is 5'-gggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:96 is 5'-ggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:97 is 5'-gg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:98 is 5'-g gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:98 is 5'-g gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEO ID NO:99 is 5'-

5

10

15

20

25

30

gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:100 is 5'-aaaagttic tticagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:101 is 5'-aaagittc tticagagag ttaagitatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agigigigig tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:102 is 5'-aagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:103 is 5'-agttic titcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:104 is 5'-gtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEO ID NO:105 is 5'-tttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:106 is 5'-ttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:107 is 5'-tc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:108 is 5'-c tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:109 is 5'-tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:110 is 5'-ttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:111 is 5'-tcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:112 is 5'-cagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEO ID NO:113 is 5'-agagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEO ID NO:114 is 5'-gagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:115 is 5'-agag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata

5

10

15

20

25

30

cacacatata atggaagcaa taagccat-3'; SEQ ID NO:116 is 5'-gag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:117 is 5'-ag ttaagttatt ttatatata aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:118 is 5'-g ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:119 is 5'-ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3': SEO ID NO:120 is 5'-taagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:121 is 5'-aagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:122 is 5'-agttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:123 is 5'-gttatt ttatatatat aatatatat aatatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:124 is 5'-ttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:125 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagcca-3'; SEQ ID NO:126 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagcc-3'; SEQ ID NO:127 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagc-3'; SEQ ID NO:128 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taag-3'; SEQ ID NO:129 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taa-3'; SEQ ID NO:130 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa ta-3'; SEQ ID NO:131 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa t-3'; SEQ ID NO:132 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata



5

10

15

20

25

30

atggaagcaa -3'; SEQ ID NO:133 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagca-3'; SEQ ID NO:134 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagc-3'; SEQ ID NO:135 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaag-3'; SEQ ID NO:136 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaa-3'; SEQ ID NO:137 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atgga-3'; SEQ ID NO:138 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atgg-3'; SEQ ID NO:139 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atg-3'; SEQ ID NO:140 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata at-3'; SEQ ID NO:141 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata a-3'; SEQ ID NO:142 is 5'-ttgggg gaaaagtttc tticagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata-3'; SEQ ID NO:143 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatat-3'; and SEQ ID NO:144 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacata-3'.

The nucleotide sequence of SEQ ID NOs:1 and 3, portions, homologs and antisense sequences thereof may be synthesized by synthetic chemistry techniques which are commercially available and well known in the art [see Caruthers MH et al., (1980) Nuc. Acids Res. Symp. Ser. 215-223; Horn T. et al., (1980) Nuc. Acids Res. Symp. Ser. 225-232]. Additionally, fragments of SEQ ID NOs:1 and 3 can be made by treatment of SEQ ID NOs:1 and 3 with restriction enzymes followed by purification of the fragments by gel electrophoresis. Alternatively, sequences may also be produced using the polymerase chain reaction (PCR) as described by Mullis [U.S. Patent Nos. 4,683,195, 4,683,202 and 4,965,188, all of which are hereby incorporated by reference]. SEQ ID NOs:1 and 3, portions, homologs and antisense sequences thereof may be ligated to each other or to heterologous nucleic acid sequences using methods well known in the art.

5

10

15

20

25

30

The nucleotide sequence of synthesized sequences may be confirmed using commercially available kits as well as using methods well known in the art which utilize enzymes such as the Klenow fragment of DNA polymerase I, Sequenase<sup>®</sup>, *Taq* DNA polymerase, or thermostable T7 polymerase. Capillary electrophoresis may also be used to analyze the size and confirm the nucleotide sequence of the products of nucleic acid synthesis, restriction enzyme digestion or PCR amplification.

It is readily appreciated by those in the art that the sequences of the present invention may be used in a variety of ways. For example, the nucleic acid sequences of the invention and portions thereof can be used as probes for the detection and isolation of functional homologs of AE5' and AE3', amplification of homologous nucleotide sequences, age-related and/or liver-specific expression of a nucleotide sequence of interest in an animal, gene therapy, and reducing factor IX levels in an animal.

ii. Regulatory Nucleic Acid Sequences From The hPC Gene

The invention provides regulatory nucleic acid sequences which are derived from the hPC gene. The presence of these sequences and their surprising properties was fortuitously discovered by the inventors during their investigation of the universality of the function of regulatory sequences from the hFIX gene. In particular, when using what they believed would be "control" constructs which contained different portions of the sequence upstream of nt +1 in the hPC gene, the inventors discovered that whereas transgenic animals containing the -1462hPCm1 construct exhibited relatively constant and relatively high levels (from

about 100 to about 3000 ng/ml) of human protein C over time (Figure 17A), in dramatic contrast, transgenic animals containing the -82hPCml construct exhibited relatively low levels (from about 5 to about 40 ng/ml at 1 month of age) (Figure 17B) of human protein C which declined at a precipitous rate over time (Example 12). Indeed, by the age of 5 months, human protein C levels were undetectable in all transgenic animals harboring the -82hPCml construct. These results demonstrated to the inventors that the nucleotide sequence from nt -1462 to nt -83 of the human protein C gene exhibits age-related regulatory activity (as evidenced by age-stable expression over time) and regulatory activity (as evidenced by the relatively higher levels of expression as compared to the levels in the absence of the nucleotide sequence from nt -1462 to nt -83) of operably linked sequences of interest.

5

10

15

20

25

30

The term "regulatory activity" when made herein in reference to a nucleic acid sequence refers to the ability of the nucleic acid sequence to alter the level of transcription into mRNA and/or the synthesis of a polypeptide encoded by a nucleotide sequence of interest which is operably linked to a promoter sequence, as compared to the level of transcription into mRNA and/or synthesis of the polypeptide encoded by the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has regulatory activity. In contrast to "age-related regulatory activity," the term "regulatory activity" relates to alteration in the level of transcription or protein synthesis at a single time point (rather than over a period of time) in the life cycle of a cell, tissue, or organism. In a preferred embodiment, the level of transcription into mRNA and/or synthesis of polypeptide is increased from 2-fold to at least 10,000-fold, preferably from 2-fold to 10,000 fold, more preferably from 2-fold to 1,000-fold, and most preferably, from 2-fold to 500-fold, when compared to the level of transcription into mRNA and/or synthesis of the polypeptide encoded by the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has regulatory activity. Data presented herein demonstrates that the inclusion of the human protein C (hPC) nucleotide sequence from nt -1462 to nt -83 resulted in expression of from about 100 to about 3000 ng/ml of human protein C (Figure 17A) as compared to expression of from about 5 to about 40 ng/ml of human protein C (Figure 17B) in transgenic mice at one-month of age. The increased level of expression was even more dramatic when the transgenic animals were 5 months-old; while the levels of hPC were undetectable in the



absence of the sequence from nt -1462 to nt -83, they were from about 100 to about 3000 ng/ml in the presence of this sequence.

5

10

15

20

25

30

In one embodiment, the invention provides the nucleic acid sequence (SEO ID NO:85) from -1462 to +1 of the hPC gene (Figure 14; 5'-GAATTCTGTA AGCATTTCCT ATGTGTACCT GCCCCTGGGC AAGGTGGGCC TGACTTGTTA GAGTGTTAGA GTTTTACCCT GTTCCTCTAG GAGGGCCTGG TACCACCACA GCCCAGCATG GTGTGGTGCC TCAGCAGGAG GCATCTGGTT ACAATCAACA CAAGCTGTTC CAGCCAATTT AAAGAAACTT CAGGAGGAAT AGGGTTTTAG GAGGGCATGG GGACCCTCCT GCACCCGAAG CCAGGATGTG CCACCAATCA TAAGGAGGCA GGGGCCTCCT TCCGCTGCTC CCTGGGACTC TCTAGGTGTC CGTGGCCTCA GCCCCCTCT GCACACCTGC ATCTTCCTTC TCATCAGCTT CCTCTGCTTT AAGCGTAAAC ATGGATGCCC AGGACCTGGC CTCAATCTTC CGAGTCTGGT ACTTATGGTG TACTGACAGT GTGAGACCCT ACTCCTCTGA TCAATCCCCT GGGTTGGTGA CTTCCCTGTG CAATCAATGG AAGCCAGCGA GGCAGGGTCA CATGCCCCGT TTAGAGGTGC AGACTTGGAG AAGGAACGTG GGCAAGTCTT CCCAGGAACA GGTAGGGCAG GGAGGAAAGG GGGGCATCTC TGGTGCAGCC CGGTTCGGAG CAGGAAGACG CTTAATAAAT GCTGATAGAC TGCAGGACAC AGGCAAAGGT GCTGAGCTGG ACCCTTTATT TCTGCCCTTC TCCCTTCTGG CACCCGGCC AGGAAATTGC TGCAGCCTTT CTGGAATCCC GTTCATTTTT CTTACTGGTC CACAAAAGGG GCCAAATGGA AGCAGCAAGA CCTGAGTTCA AATTAAATCT GCCAACTACC AGCTCAGTGA ATCTGGGCGA GTAACACAAA ACTTGAGTGT CCTTACCTGA AAAATAGAGG TTAGAGGGAT GCTATGTGCC ATTGTGTGT TGTGTTGGGG GTGGGGATTG GGGGTGATTT GTGAGCAATT GGAGGTGAGG GTGGAGCCCA GTGCCCAGCA CCTATGCACT GGGGACCCAA AAAGGAGCAT CTTCTCATGA TTTTATGTAT CAGAAATTGG GATGGCATGT CATTGGGACA GCGTCTTTTT TCTTGTATGG TGGCACATAA ATACATGTGT CTTATAATTA ATGGTATTTT AGATTTGACG AAATATGGAA TATTACCTGT TGTGCTGATC TTGGGCAAAC TATAATATCT CTGGGCAAAA ATGTCCCCAT CTGAAAAACA GGGACAACGT TCCTCCCTCA GCCAGCCACT ATGGGGCTAA AATGAGACCA CATCTGTCAA GGGTTTTGCC CTCACCTCCC TCCCTGCTGG ATGGCATCCT TGGTAGGCAG AGGTGGGCTT CGGGCAGAAC AAGCCGTGCT GAGCTAGGAC CAGGAGTGCT AGTGCCACTG TTTGTCTATG GAGAGGGAGG

CCTCAGTGCT GAGGGCCAAG CAAATATTTG TGGTTATGGA TTA-3'). This sequence was successfully used to express hPC in transgenic mice in an age-related manner (i.e., at relatively constant levels over time; Figure 17A), and at relatively high levels (e.g., as compared to expression levels in the presence of sequences from nt -1462 to nt -82; Figure 17B).

5

10

15

20

25

30

The invention also provides the nucleotide sequence (SEQ ID NO:92) from -1462 to -83 of the hPC gene (Figure 14). The finding that SEQ ID NO:92 exhibited age-related regulatory activity and regulatory activity as discussed in Example 11, infra, was surprising because it was contrary to the results which had previously been reported by Miao et al. [Miao et al. (1996) J. Biol. Chem. 16:9587-9594] when using a heterologous reporter gene, chloramphenicol acetyltranferase (CAT), under the transcriptional control of varying lengths of the protein C 5'-end sequences. In particular, Miao et al. found that construct pPC-1528 which contained nucleotides -1462 to +1 resulted in substantially reduced CAT in vitro activity as compared to construct pPC-82-66 which contained nucleotides -82 to +1. From this, Miao et al: concluded that there is an element with silencer activity in the region between -1462 and -82. In contrast, data presented herein demonstrates that nucleotides -1462 to +1 and -82 to +1 resulted in similar in vitro activities of hPC (Example 11, Figure 16). Indeed, the inventors' data disclosed in relation to hPC's age-related regulatory sequences which are upstream of the hPC coding sequences when used to regulate expression of hPC is consistent with the inventor's observation (Figure 1) in relation to hFIX's agerelated regulatory AE5' sequences when used to regulate expression of hFIX. In particular, the inventors have observed that hFIX's AE3' and AE3" sequences resulted in moderate suppression in in vitro transient expression assays when using hFIX and hPC, respectively (Figures 1 and 16).

Further contemplated to be within the scope of this invention are portions of the SEQ ID NOS:85 and 92. In a particularly preferred embodiment, these portions contain one or both of the first PEA-3 element (SEQ ID NO:89) [7-bp long; 5'-GAGGAAA-3', from -871 to -865 of the hPC gene of Figure 14] and the second PEA-3 element (SEQ ID NO:90) [7-bp long; 5'-CAGGAAG-3', from -832 to -826 of the hPC gene of Figure 14].

Exemplary portions of SEQ ID NOs:85 and 92 which contain both the first and second PEA-3 elements include, but are not limited to, the nucleotide sequence (SEQ ID NO:88) from -1462 to -802 of the hPC gene, which is embodied in plasmid -849hPCm1

(Figure 15, Example 10). Further examples include the nucleotide sequence (SEQ ID NO:145) from -1462 to -804, (SEQ ID NO:146) from -1462 to -805, (SEQ ID NO:147) from -1462 to -806, (SEQ ID NO:148) from -1462 to -807, (SEQ ID NO:149) from -1462 to -808, (SEQ ID NO:150) from -1462 to -809, (SEQ ID NO:151) from -1462 to -810, (SEQ ID NO:152) from -1462 to -811, (SEQ ID NO:153) from -1462 to -812, (SEQ ID NO:154) from -1462 to -813, (SEQ ID NO:155) from -1462 to -814, (SEQ ID NO:156) from -1462 to -815, (SEQ ID NO:157) from -1462 to -816, (SEQ ID NO:158) from -1462 to -817, (SEQ ID NO:159) from -1462 to -818, (SEQ ID NO:160) from -1462 to -819, (SEQ ID NO:161) from -1462 to -820, (SEQ ID NO:162) from -1462 to -821, (SEQ ID NO:163) from -1462 to -822, (SEQ ID NO:164) from -1462 to -823, (SEQ ID NO:165) from -1462 to -824, (SEQ ID NO:166) from -1462 to -825, (SEQ ID NO:167) from -1462 to -826, (SEQ ID NO:168) from -1452 to -803, (SEQ ID NO:169) from -1442 to -803, (SEQ ID NO:170) from -1412 to -803, (SEQ ID NO:171) from -1102 to -803, (SEQ ID NO:172) from -902 to -803, and (SEQ ID NO:173) from -873 to -803.

5

10

15

20

25

30

Portions of SEQ ID NOs:85 and 92 which contain only the first PEA-3 element include the nucleotide sequence (SEQ ID NO:87) from -1462 to -849 of the hPC gene, which is embodied in plasmid -802hPCm1 (Figure 15, Example 10). Additional exemplary portions include, but are not limited to the nucleotide sequence (SEQ ID NO:174) from -1462 to -850, (SEQ ID NO:175) from -1462 to -851, (SEQ ID NO:176) from -1462 to -852, (SEQ ID NO:177) from -1462 to -853, (SEQ ID NO:178) from -1462 to -854, (SEQ ID NO:179) from -1462 to -855, (SEQ ID NO:180) from -1462 to -856, (SEQ ID NO:181) from -1462 to -857, (SEQ ID NO:182) from -1462 to -858, (SEQ ID NO:183) from -1462 to -859, (SEQ ID NO:184) from -1462 to -860, (SEQ ID NO:185) from -1462 to -861, (SEQ ID NO:186) from -1462 to -862, (SEQ ID NO:187) from -1462 to -863, (SEQ ID NO:190) from -1362 to -865, (SEQ ID NO:191) from -1262 to -865, (SEQ ID NO:192) from -1162 to -865, (SEQ ID NO:193) from -1062 to -865, (SEQ ID NO:194) from -962 to -865, and (SEQ ID NO:195) from -872 to -865.

Examples of portions of SEQ ID NOs:85 and 92 which contain only the second PEA-3 element include, but are not limited to, the nucleotide sequence (SEQ ID NO:196) from -863 to -83, (SEQ ID NO:197) from -853 to -83, (SEQ ID NO:198) from -843 to -83, (SEQ ID NO:199) from -833 to -83, (SEQ ID NO:200) from -832 to -83, (SEQ ID NO:201) from

-863 to -183, (SEQ ID NO:202) from -863 to -283, (SEQ ID NO:203) from -863 to -383, (SEQ ID NO:204) from -863 to -483, (SEQ ID NO:205) from -863 to -583, (SEQ ID NO:206) from -863 to -683, (SEQ ID NO:207) from -863 to -783, and (SEQ ID NO:208) from -863 to -826.

In a particularly preferred embodiment, the portion of SEQ ID NOs:85 and 92 is selected from the first PEA-3 element (SEQ ID NO:89) and the second PEA-3 element (SEQ ID NO:90). It is the inventor's view that the first and/or second PEA-3 elements within SEQ ID NOs:85 and 92 are responsible for the observed age-related regulatory activity and regulatory activity of SEQ ID NOs:85 and 92.

10

15

20

25

30

5

# B. Using Probes To Identify And Isolate Homologs Of AE5', AE3', and Of hPC-Derived Regulatory Sequences

The invention provided herein is not limited to SEQ ID NO:1, 3, 85 and 92, homologs and portions thereof having age-related regulatory activity, but includes sequences having no age-related regulatory activity (i.e., non-functional homologs and non-functional portions of homologs). The use of such sequences may be desirable, for example, where a portion of SEQ ID NOs:1, 3, 85, and 92 is used as a probe to detect the presence of SEQ ID NOs:1, 3, 85 and 92, respectively, or of portions thereof in a sample.

As used herein, the term "probe" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to a nucleotide sequence of interest. A probe may be single-stranded or double-stranded. It is contemplated that any probe used in the present invention will be labelled with any "reporter molecule," so that it is detectable in any detection system including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, calorimetric, gravimetric, magnetic, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The probes provided herein are useful in the detection, identification and isolation of, for example, sequences such as those listed as SEQ ID NOs:1, 3, 85 and 92 as well as of homologs thereof. Preferred probes are of sufficient length (e.g., from about 9 nucleotides to about 20 nucleotides or more in length) such that high stringency hybridization may be

employed. In one embodiment, probes from 20 to 50 nucleotide bases in length are employed.

# C. Using Primers to Amplify At Least A Portion Of AE5', AE3', and Of hPC-Derived Regulatory Sequences

5

10

15

20

25

30

The invention provided herein is not limited to SEQ ID NOs:1 and 3, homologs and portions thereof having age-related regulatory activity, but includes sequences having no agerelated regulatory activity. This may be desirable, for example, where a portion of the nucleic acid sequences set forth as SEQ ID NOs:1 and 3 is used as a primer for the amplification of nucleic acid sequences by, for example, polymerase chain reactions (PCR) or reverse transcription-polymerase chain reactions (RT-PCR). The term "amplification" is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction technologies well known in the art [Dieffenbach CW and GS Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY]. As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis disclosed in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,965,188, all of which are hereby incorporated by reference, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter

"PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are the to be "PCR amplified."

5

10

15

20

25

30

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; and/or incorporation of <sup>32</sup>P-labeled deoxyribonucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any nucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. Amplified target sequences may be used to obtain segments of DNA (e.g., genes) for the construction of targeting vectors, transgenes, etc.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (i.e., in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long (e.g., from about 9 nucleotides to about 20 nucleotides or more in length) to prime the synthesis of extension products in the presence of the inducing agent. Suitable lengths of the primers may be empirically determined and depend on factors such as temperature, source of primer and the use of the method. In one embodiment, the present invention employs primers from 20 to 50 nucleotide bases in length.

The primers contemplated by the invention are useful in, for example, identifying sequences which are homologous to AE5', AE3', and regulatory sequences derived from hPC, in mammals, yeast, bacteria, and in other organisms.

# D. Methods For Regulating Gene Expression

5

10

15

20

25

30

The present invention provides methods for regulating expression of a nucleotide sequence of interest over a period of time in a cell or multicellular organism. Specifically, gene expression is preferably regulated in a multicellular organism. In one embodiment, expression of a nucleotide sequence of interest is stabilized such that the level of mRNA and/or protein encoded by the nucleotide sequence of interest remains relatively unchanged at different times during the life of the organism. In an alternative embodiment, expression of a nucleotide sequence of interest is increased. Increased expression means that the level of mRNA and/or protein encoded by the nucleotide sequence of interest at a given time point is greater than the level of mRNA and/or protein, respectively, at an earlier time point during the life of the organism or cell. Alternatively, increased expression means that the level of mRNA and/or protein encoded by the nucleotide sequence of interest is greater than the level of mRNA and/or protein, respectively, at the same time point in the life of the organism or cell as compared to the level of mRNA and/or protein when expressed in the absence of the sequences of the invention.

In one embodiment, regulating expression of a nucleotide sequence of interest over a period of time is accomplished by introducing into a host cell a vector that contains a nucleotide sequence of interest operably linked to a promoter sequence and to sequences provided herein which have age-related regulatory activity. The transfected host cell is allowed to develop into a transgenic animal in which the nucleotide sequence of interest is expressed in at least one tissue. These steps are further described below for specific embodiments.

#### 1. Expression Constructs

In one embodiment of the methods of the invention for regulating expression of a nucleotide sequence of interest in an age-related manner and/or to liver tissue, a vector is constructed in which the nucleic acid sequences of the invention (e.g., AE5' alone, AE3' alone, or a combination of AE5' and AE3') are operably linked to a promoter sequence and to a nucleotide sequence of interest. In one embodiment, the nucleotide sequence of interest is the coding region of the hFIX gene (Example 1). In another embodiment the nucleotide sequence of interest is the coding region of the protein C gene (Example 7).

The invention is not limited to coding sequences of the hFIX gene or protein C gene. Rather, any nucleotide acid sequence whose expression is desired to be regulated by sequences provided herein are contemplated to be within the scope of this invention. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes which encode a protein [e.g., reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factor genes, activator protein 1 gene, activator protein 2 gene, Sp1 gene, etc.]. In one preferred embodiment, the structural gene is the human α1-antitrypsin gene (Figure 10) (SEQ ID NO:42) which encodes a plasma proteinase inhibitor used for treating emphysema. In another preferred embodiment, the structural gene is one encoding the human antithrombin III (Figure 11) (SEQ ID NO:43) which is a plasma proteinase inhibitor for activated blood coagulation factors and whose activity is increased by heparin. In yet another preferred embodiment, the structural gene is the gene encoding the PEA-3 protein (Figure 9) (SEQ ID NO:47) and/or its related protein, which has been shown to bind specifically to homologs of the PEA-3 nucleotide sequence (SEQ ID NO:2) disclosed herein.

5

10

15

20

25

30

The invention is not limited to using a single nucleotide sequence of interest in operable combination with the sequences of the invention. Rather, a plurality (i.e., more than one) of nucleotide sequences of interest may be ligated in tandem such that their expression is regulated by the regulatory sequences of the invention. A plurality of coding sequences may be desirable, for example, where it is useful to express a transcription product of more than one gene to permit interaction of these transcriptional products. Alternatively, a plurality of coding sequences may be desirable where one of the gene sequences is a reporter gene sequence. For example, it may be advantageous to place a coding sequence of a reporter gene in tandem with the coding sequence of a gene of interest such that expression of the coding region of both the reporter gene and the gene of interest is regulated by the sequences of the invention. Expression of the reporter gene usually correlates with expression of the gene of interest. Examples of reporter gene sequences include the sequences encoding the enzymes β-galactosidase and luciferase. Fusion genes may also be desirable to facilitate purification of the expressed protein. For example, the heterologous sequence which encodes protein A allows purification of the fusion protein on immobilized immunoglobulin. Other affinity traps are well known in the art and can be utilized to advantage in purifying the expressed fusion protein. For example, pGEX vectors (Promega, Madison WI) may be used to express the polypeptides of interest as a fusion protein with

glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Other fusion polypeptides useful in the purification of proteins of interest are commercially available, including histidine tails (which bind to Ni<sup>2+</sup>), biotin (which binds to streptavidin), and maltose-binding protein (MBP) (which binds to amylose). Proteins made in such systems may be designed to include heparin, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released at will from the heterologous polypeptide moiety to which it is fused.

5

10

15

20

25

30

One of skill in the art would understand that where a plurality of nucleotide sequences of interest is operably linked to sequences of the present invention, the nucleotide sequences of interest may be either contiguous or separated by intervening polynucleotide sequences, so long as the nucleic acid sequences of interest are operably linked to the promoter sequence, and so long as the sequences of the invention are operably linked to the promoter sequence.

While specific preferred embodiments used herein disclose the use of the hFIX promoter and the CMV promoter, it is not intended that the invention be limited to the type or source of the promoter sequence which is operably linked to the sequences of the invention. Any promoter whose activity is desired to be regulated by the sequences provided herein is contemplated to be within the scope of the invention. Exemplary promoters include the tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter (can be isolated from vector plasmid pRc/RSV from Invitrogen), retrovirus LTR promoter (can be isolated from vector plasmid pLXSN from Clontech) SV40 promoter (located between positions +3530 to +3192 in vector plasmid pCR3 from Invitrogen), PEPCK promoter, MT promoter, SRa promoter, P450 family promoters, GAL7 promoter, T, promoter having the 23-bp sequence (SEQ ID NO:44) 5'-TAATACGACTCACTATAGGGCGA-3', T, promoter having the 24-bp sequence (SEQ ID NO:45) 5'-TTATTAACCCTCACTAAAGGGAAG -3', SP6 promoter having the 23-bp sequence (SEQ ID NO:46) 5'-ATTTAGGTGACACTATAGAATAC -3', and K11 promoter. The T, promoter, T, promoter, SP6 promoter and K11 promoter have been described in U.S Patent No.

Nor is the invention intended to be limited to the use of a single promoter. For example, chimeric promoters (i.e., two or more promoters which are derived from at least

5,591,601, the entire contents of which are incorporated by reference.

one gene) are expressly contemplated to be within the scope of the invention. Such chimeric promoters may be desirable where, for example, chimeric promoters result in increased levels of expression of an operably linked downstream coding sequence. Chimeric promoters are known in the art and include, for example, the double *Tet* promoter [Kistner et al. (1996) Proc. Natl. Acad. Sci. USA 93:10933-10938], and the U1 snRNA promoter-CMV promoter/enhancer [Bartlett et al. (1996) Proc. Natl. Acad. Sci. USA 93:8852-8857].

5

10

15

20

25

30

Expression vectors in which expression of a nucleic acid sequence of interest is regulated by sequences of the invention may be constructed using the teachings of the present invention in conjunction with techniques well known in the art. [Sambrook et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY; Ausubel et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY]. Briefly, the nucleic acid sequence of interest is placed in operable combination with a promoter sequence and sequences of the invention in the presence of transcription and translation regulatory sequences, including initiation signals such as a start codon (i.e., ATG), enhancers, and transcription termination signals. The ATG initiation codon must be in the correct reading frame to ensure translation of the entire heterologous nucleotide sequence. Transcription termination signals are placed downstream of the heterologous nucleic acid sequence and include polyadenylation sequences which are exemplified by, but not limited to, SV40 poly-A sequence, hINV poly-A sequence, or bovine growth hormone poly-A sequence, etc.

Other regulatory sequences which may affect RNA stability as well as enhancers (i.e., a sequence which when activated results in an increase in the basal rate of transcription of a gene) and silencers (i.e., a sequence involved in reducing expression of a gene) may also be included. These regulatory sequences may be relatively position-insensitive, i.e., the regulatory element will function correctly even if positioned differently in relation to the heterologous nucleotide sequence in the construct as compared to its position in relation to the corresponding heterologous nucleotide sequence in the genome. For example, an enhancer may be located at different distances from the promoter sequence, in a different orientation, and/or in a different linear order. Thus, an enhancer that is located 3' to a promoter sequence in germline configuration might be located 5' to the promoter sequence in the construct.

#### 2. Host Cells

5

10

15

20

25

30

Host cells are transformed with expression vectors which contain the sequences of the invention in operable combination with a nucleic acid sequence of interest using methods known in the art. The term "transformation" as used herein refers to the introduction of a transgene into a cell. The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations.

The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations. A transgene may be an "endogenous DNA sequence," or a "heterologous DNA sequence." The term "endogenous DNA sequence" refers to a nucleotide sequence which is naturally found in the cell into which it is introduced so long as it does not contain some modification (e.g., a point mutation, the presence of a selectable marker gene, etc.) relative to the naturally-occurring sequence. The terms "heterologous DNA sequence" and "foreign DNA sequence" refer to a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Heterologous DNA also includes an endogenous DNA sequence which contains some modification (e.g., a point mutation, the presence of a selectable marker gene, etc.) relative to the naturally-occurring gene. Generally, although not necessarily, heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, selectable marker proteins (e.g., proteins which confer drug resistance), etc.

Transformation may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, retroviral infection, biolistics (*i.e.*, particle bombardment) and the like.

Transformation of a cell may be stable or transient. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) which detects the presence of a polypeptide encoded by one

5

10

15

20

25

30

or more of the transgenes. Alternatively, transient transformation may be detected by detecting the activity of the protein encoded by the transgene. For example, the activity of  $\beta$ -glucuronidase (GUS) which is encoded by the *uid A* gene may be detected using either a histochemical assay of GUS enzyme activity by staining with X-gluc which gives a blue precipitate in the presence of the GUS enzyme, or a chemiluminescent assay using the GUS-Light kit (Tropix). The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction (PCR) of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell which has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene.

Suitable host cells include bacterial, yeast, plant, insect, and mammalian cells. In one embodiment the host cell is mammalian. In a preferred embodiment, the mammalian host cell is a mouse fertilized egg cell. In an alternative embodiment, the mammalian host cell is a HepG2 cell (ATCC number HB8065), a fibroblast cell (e.g., ATCC number CCL 110), a myoblast cell (e.g., Clonetics, catalog # SkMC), and an endothelial cell (e.g., human umbilical cord endothelial cells; ATCC number CRL 1730).

In one embodiment, the host cell is transformed both with an expression vector which contains the sequences of the invention in operable combination with the nucleic acid sequences of interest, as well as with an expression vector which expresses the PEA-3 protein (Example 6). Such co-transformation may be desirable, for example, where expression of the nucleotide sequence of interest is regulated by AE5' or portions or homologs thereof which contain homologs of the PEA-3 nucleotide sequence to which the PEA-3 protein binds. In one embodiment, expression of the PEA-3 protein is under the control of the LTR promoter of the Moloney murine leukemia virus (MoLV) which is capable of driving expression of operably linked genes in several cell types. Transient

expression assays are suitable for determining the relative promoter activities in expressing desirable PEA-3 protein levels.

5

10

15

20

25

30

Any number of selection systems may be used to recover transfected cells. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy I et al. (1980) Cell 22:817-23) genes which can be employed in the or aprt cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate [Wigler M et al., (1980) Proc Natl Acad Sci 77:3567-70]; npt, which confers resistance to the aminoglycosides neomycin and G-418 [Colbere-Garapin F et al., (1981) J. Mol. Biol. 150:1-14] and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine [Hartman SC and RC Mulligan (1988) Proc Natl Acad Sci 85:8047-51]. Recently, the use of a reporter gene system which expresses visible markers has gained popularity with such markers as β-glucuronidase and its substrate (GUS), luciferase and its substrate (luciferin), and β-galactosidase and its substrate (X-Gal) being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system [Rhodes CA et al. (1995) Methods Mol Biol 55:121-131].

The presence or expression of the reporter gene usually indicates the presence or expression, respectively, of the tandem heterologous nucleic acid sequence as well. However, it is preferred that the presence and expression of the desired heterologous nucleic acid sequence be confirmed. This is accomplished by procedures known in the art which include DNA-DNA or DNA-RNA hybridization or amplification using probes, or fragments of the heterologous nucleic acid sequence. For example, Fluorescent In Situ Hybridization (FISH) can be used to detect the heterologous nucleic acid sequence in cells. Several guides to FISH techniques are available, e.g., Gall et al. Meth. Enzymol. 21:470-480 (1981); Angerer et al., in "Genetic Engineering: Principles and Methods," Setlow & Hollaender, Eds. Vol. 7 pp. 43-65, Plenum Press, New York (1985). Alternatively, DNA or RNA can be isolated from cells for detection of the transgene by Southern or Northern hybridization or by amplification based assays. Nucleic acid amplification based assays involve the use of

oligonucleotides or oligomers based on the nucleotide sequence of interest in order to detect cells and tissues which contain the DNA or RNA encoding the transgene of interest. As used herein, the terms "oligonucleotides" and "oligomers" refer to a nucleic acid sequence of at least about five (5) contiguous nucleotide residues and as many as about sixty (60) nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplimer. Standard PCR methods useful in the present invention are described by Innis *et al.* (Eds.), "PCR Protocols: A Guide to Methods and Applications," Academic Press, San Diego (1990).

5

10

15

20

25

30

Yet another alternative for the detection of heterologous nucleic acid sequences includes detecting the polypeptide product of transcription of the heterologous nucleotide sequence. A variety of protocols which employ polyclonal or monoclonal antibodies specific for the protein product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A competitive binding assay may also be used. Alternatively, a two-site, monoclonal-based immunoassay which utilizes monoclonal antibodies that are reactive to two non-interfering epitopes on the protein of interest may be employed. These and other assays are described in, among other places, Hampton R et al. (1990), Serological Methods a Laboratory Manual, APS Press, St Paul MN), and Maddox DE et al. (1983), J. Exp. Med. 158:1211.

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting related sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the nucleotide sequence of interest, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides. A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like.

Host cells transformed with expression vectors containing the sequences provided herein are useful for age-related expression of recombinant proteins of interest. Host cells transformed with expression vectors containing the invention's sequences may be part of a tissue or organ of a living animal. A "living animal" as used herein refers to any multicellular animal (e.g., humans, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc.) into whose cells the sequences provided herein may be introduced. Where the host cells (e.g., fertilized egg cells) are capable of generating a multicellular organism, these cells when transformed with expression vectors containing the sequences of the invention are useful in generating transgenic animals which exhibit age-related and/or liver-specific expression of nucleotide sequences of interest.

# · 3. Transgenic Animals

5

10

15

20

25

30

The present invention provides transgenic non-human animals which express a nucleotide sequence of interest in an age-related manner. These animals provide useful models for diseases (e.g., thrombosis, cardiovascular diseases, diabetes, Alzheimer's disease, cancer, osteoporosis, osteoarthritis, Parkinson's disease, dementia) which are associated with increasing age, as well a for screening candidate therapeutic agents against such diseases. These transgenic animals are also useful in studies of normal phenomena, such as ageing, gene regulation, etc. In one embodiment, the invention discloses transgenic mice which express in an age-related manner the exemplary hFIX coding sequence under the control of AE5' and/or AE3' (Example 3).

The term "age-related manner" when made in reference to the expression of a nucleotide sequence of interest is a relative term which refers to an increase over a period of time in the quantity of mRNA and/or protein encoded by the nucleotide sequence of interest when the nucleotide sequence of interest is operably linked to a promoter and to a nucleic acid sequence which has age-related regulatory activity, as compared to the quantity of mRNA and/or protein, respectively, encoded by the nucleotide sequence of interest when the nucleotide sequence of interest is operably linked to the promoter in the absence of the nucleic acid sequence which has age-related regulatory activity. Thus, the term "age-related" when made in reference to expression of a nucleotide sequence of interest by a transgenic

animal means that the transgenic animal expresses the nucleotide sequence of interest in an age-related manner.

For example, in one embodiment, the invention demonstrates that hFIX is expressed in an age-related manner in transgenic mice which harbor a transgene (-416FIXm1/1.4) (Figure 2B) which contains hFIX under the control of the hFIX promoter and the regulatory control of AE3' as compared to expression of hFIX in transgenic mice which harbor a transgene (-416FIXm1) (Figure 2A) in which hFIX is under the control of the hFIX promoter in the absence of AE3'. While transgenic mice harboring the -416FIXm1/1.4 construct showed decreasing hFIX activity levels over a period of time (e.g., from 1 to 9 months of age), this decrease was less than the decrease in hFIX activity levels which was observed in transgenic mice harboring the -416FIXm1 construct over the same period of time.

5

10

15

20

25

30

In another embodiment, the invention discloses that hFIX is expressed in an age-related manner in transgenic mice which harbor transgenes (-802FIXm1, -2231FIXm1, and -416FIXm1/AE5') (Figures 4A, C and E) each of which contains hFIX under the control of the hFIX promoter and the regulatory control of AE5' as compared to expression of hFIX in transgenic mice which harbor a transgenes (-416FIXm1 and -770FIXm1) (Figures 2A and E) in which hFIX is under the control of the hFIX promoter in the absence of AE5'.

Transgenic mice harboring each of the -802FIXm1, -2231FIXm1, and -416FIXm1/AE5' constructs showed relatively unchanged hFIX activity levels over a period of time (e.g., from 1 to 7 months of age) while transgenic mice harboring either the -416FIXm1 or -770FIXm1 construct showed decreasing hFIX activity levels over the same time period.

In an additional embodiment, the invention shows that hFIX is expressed in an agerelated manner in transgenic mice which harbor transgenes (-802FIXm1/1.4 and
-2231FIXm1/1.4) (Figures 4B and D) each of which contains hFIX under the control of the
hFIX promoter and the regulatory control of both AE3' and AE5' as compared to expression
of hFIX in transgenic mice which harbor a transgene (-770FIXm1) (Figure 2E) in which
hFIX is under the control of the hFIX promoter in the absence of both AE3' and AE5'.

Transgenic mice harboring either the -802FIXm1/1.4 or the -2231FIXm1/1.4 construct
showed increasing levels of hFIX activity over a period of time (e.g., 1 to 3 months of age)
as compared to decreasing hFIX activity levels over the same period of time in transgenic
mice harboring the -770FIXm1 construct.

The present invention also provides transgenic non-human animals which express a nucleotide sequence of interest in a liver-specific manner. These animals are useful for targeting expression of a nucleotide sequence of interest to the liver. Examples of nucleotide sequences of interest are those which encode blood coagulation factors (e.g., factor VIII, factor VII, factor X and prothrombin) whose deficiency is known to play a role in abnormal bleeding disorders. Other examples of nucleotide sequences of interest include those which encode blood coagulation regulators and/or inhibitors (e.g., protein C, antithrombin III, and tissue factor pathway inhibitor [TFPI]) whose deficiency results in thrombosis, α1-antitrypsin whose deficiency results in emhysima, and LDL-receptor whose deficiency results in hypercholestrolemia.

5

10

15

20

25

30

Yet other examples of a nucleotide sequence of interest include those encoding enzymes involved in specific metabolic defects (e.g., urea cycle enzymes, especially ornithine transcarbamylase, argininosuccinate synthase, and carbamyl phosphate synthase); receptors (e.g., LDL receptor); toxins; thymidine kinase to ablate specific cells or tissues; ion channels (e.g., chloride channel of cystic fibrosis); membrane transporters (e.g., glucose transporter); and cytoskeletal proteins (e.g., dystrophin). The nucleotide sequence of interest may be of synthetic, cDNA, or genomic origin, or a combination thereof. The nucleotide sequence of interest may be one which occurs in nature, a non-naturally occurring gene which nonetheless encodes a naturally occurring polypeptide, or a gene which encodes a recognizable mutant of such a polypeptide. It may also encode an mRNA which will be "antisense" to a DNA found or to an mRNA normally transcribed in the host cell, but which antisense RNA is not itself translatable into a protein. In one embodiment, the invention discloses transgenic mice which express in a liver-specific manner the exemplary hFIX coding sequence under the control of AE5' (Example 3).

The term "liver-specific manner" as used herein in reference to the expression of a nucleotide sequence of interest in a transgenic animal is a relative term which means that the quantity of mRNA and/or protein encoded in liver tissue by the nucleotide sequence of interest is greater than, preferably two times greater, more preferably five times greater, and most preferably ten times greater, than the quantity of mRNA and/or protein encoded by the nucleotide sequence of interest in tissues other than liver tissue of the same transgenic animal as detected by Northern blot hybridization and/or by the activity of the encoded protein as described herein. Thus, the term "liver-specific" when made in reference to expression of a

nucleotide sequence of interest by a transgenic animal means that the transgenic animal expresses the nucleotide sequence of interest in an liver-specific manner.

A first step in the generation of the transgenic animals of the invention is the introduction of a construct containing nucleic acid sequences of interest under the expression regulatory control of sequences of the invention into target cells. Several methods are available for accomplishing this, including microinjection, retroviral infection, and implantation of embryonic stem cells. These methods are discussed as follows.

# i. Microinjection Methods

5

10

15

20

25

30

Direct microinjection of expression vectors into pronuclei of fertilized eggs is the preferred, and most prevalent, technique for introducing heterologous nucleic acid sequences into the germ line. Technical aspects of the microinjection procedure and important parameters for optimizing integration of nucleic acid sequences have been previously described [Hogan *et al.*, (1986) Manipulation of the Mouse Embryo: A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Lab.].

Once the expression vector has been injected into the fertilized egg cell, the cell is implanted into the uterus of a pseudopregnant female and allowed to develop into an animal. Of the founder transgenic animals born, 70% carry the expression vector sequence in all of their cells, including the germ cells. The remaining 30% of the transgenic animals are chimeric in somatic and germ cells because integration of the expression vector sequence occurs after one or more rounds of replication. Heterozygous and homozygous animals can then be produced by interbreeding founder transgenics. This method has been successful in producing transgenic mice, sheep, pigs, rabbits and cattle [Hammer *et al.*, (1986) J. Animal Sci.:63:269; Hammer *et al.*, (1985) Nature 315:680-683].

#### ii. Retroviral Methods

Retrovirul infection of pre-implantation embryos with genetically engineered retroviruses may also be used to introduce transgenes into an animal cell. For example, blastomeres have been used as targets for retroviral infection [Jaenisch, (1976) Proc. Natl. Acad. Sci USA 73:1260-1264]. Transfection is typically achieved using a replication-defective retrovirus carrying the transgene [Jahner *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82:6927-6931; Van der Putten *et al.*, (1985) Proc. Natl. Acad Sci USA 82:6148-6152].

Transfection is obtained, for example, by culturing eight-cell embryos, from which the zona pellucida has been removed with fibroblasts which produce the virus [Van der Putten (1985), supra; Stewart et al., (1987) EMBO J. 6:383-388]. The transfected embryos are then transferred to foster mothers for continued development. Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele [Jahner et al., (1982) Nature 298:623-628]. Yet another alternative method involves intrauterine retroviral infection of the midgestation embryos [Jahner et al. (1982), supra].

The advantages of retroviral infection methods include the ease of transfection and the insertion of a single copy of the transgene, which is flanked by the retroviral long terminal repeats (LTRs), into the chromosome. However, this method is not a preferred method because most of the founders will show mosaicism since infection occurs after cell division has begun. This necessitates outbreeding to establish homozygous and heterozygous lines suitable for analysis of gene expression. More importantly, the retroviral LTR sequences may interfere with the activity of the hINV upstream sequences in directing expression of the heterologous nucleic aid sequences.

### iii. Embryonic Stem Cell Implantation

5

10

15

20

25

30

Another method of introducing transgenes into the germ line involves using embryonic stem (ES) cells as recipients of the expression vector. ES cells are pluripotent cells directly derived from the inner cell mass of blastocysts [Doetchman et al., (1988) Dev. Biol. 127:224-227], from inner cell masses [Tokunaga et al., (1989) Jpn. J. Anim. Reprod. 35:173-178], from disaggregated morulae [Eistetter, (1989) Dev. Gro. Differ. 31:275-282] or from primordial germ cells [Matsui et al., (1992) Cell 70:841-847]. Expression vectors can be introduced into ES cells using any method which is suitable for gene transfer into cells, e.g., by transfection, cell fusion, electroporation, microinjection, DNA viruses, and RNA viruses [Johnson et al., (1989) Fetal Ther. 4 (Suppl. 1):28-39].

The advantages of using ES cells include their ability to form permanent cell lines in vitro, thus providing an unlimited source of genetic material. Additionally ES cells are the most pluripotent cultured animal cells known. For example, when ES cells are injected into an intact blastocyst cavity or under the zona pellucida, at the morula stage embryo, ES cells are capable of contributing to all somatic tissues including the germ line in the resulting chimeras.

Once the expression vector has been introduced into an ES cell, the modified ES cell is then introduced back into the embryonic environment for expression and subsequent transmission to progeny animals. The most commonly used method is the injection of several ES cells into the blastocoel cavity of intact blastocysts [Bradley et al., (1984) Nature 309:225-256]. Alternatively, a clump of ES cells may be sandwiched between two eight-cell embryos [Bradley et al., (1987) in "Teratocarcinomas and Embryonic Stem Cells: A Practical Approach," Ed. Robertson E.J. (IRL, Oxford, U.K.), pp. 113-151; Nagy et al., (1990) Development 110:815-821]. Both methods result in germ line transmission at high frequency.

Target cells which contain the heterologous nucleic acid sequences are recovered, and the presence of the heterologous nucleic acid sequence in the target cells as well as in the animal is accomplished as described *supra*.

# E. Gene Therapy

5

10

15

20

25

30

The regulatory nucleic acid sequences provided herein may be used for gene therapy applications in both non-human animals as well as in humans. For example, the regulatory nucleic acid sequences of the invention may be introduced into cells using an expression vector which encodes a polypeptide sequence of interest using a variety of means known in the art to be useful both for delivery *in vivo* and *ex vivo*, including (1) recombinant retroviral transduction, (2) recombinant adenoviral vectors, (3) targeted cationic liposomes, and (4) gene transfer using biolistics, as described in the following sections.

## 1. Recombinant Retroviral Transduction

Retroviral vectors encoding polypeptides of interest may be used for the expression of the polypeptides in any desired cell, such as primary tumor cells. The transfer of polypeptides of interest using retroviruses may be made more efficient by increasing the titer of the virus encoding the polypeptides of interest and increasing the transduction efficiency. To increase the virus titer, the retroviral construct may be designed to include a selectable marker (e.g., neo gene), and cells harboring the retroviral construct are selected by growth in the presence of a suitable selective agent (e.g., G418) followed by expansion of clones producing the highest titers of virus. To improve the transduction efficiency, retrovirus are

used in combination with liposomes or poly-L-ornithine or polylysine to enhance virus uptake.

Another way to improve gene transfer efficiency using retroviruses is to increase the targeting efficiency. Many tumor cells including glioblastomas and melanomas express excess levels of the transferrin receptor. Transferrin has been used to increase the transduction efficiency of adenovirus in combination with polylysine. Several recent reports demonstrated that replacing the SU (surface) domain of the env gene of a retrovirus can increase receptor-mediated transduction efficiency. The human transferrin gene is 2097 bp long and its insertion into the SU domain of the env gene of MLV vector may not produce a stable Env product. However, since earlier studies have suggested that the modified Env fusion protein requires the native Env for stable assembly and efficient entry, co-transfection of the transferrin-env fusion gene with the native env gene may be used to produce retrovirus particles bearing a mixture of wild type and recombinant Env. The gene transfer efficiency of the new vector may be examined by transducing tumor cells expressing high levels of transferrin receptor.

#### 2. Recombinant Adenoviral Vectors

5

10

15

20

25

30

Recombinant adenoviruses can accommodate relatively large segments of foreign DNA (~7 kb), and have the advantage of a broad host cell range and high titer virus production. Adenoviruses have been used *in vivo* in rats to efficiently deliver genes to the liver and the pancreatic islets [reviewed in Becker *et al.* (1994) In *Protein Expression in Animal Cells*, Roth *et al.* eds.] and to the central nervous system [Davidson *et al.* (1993) Nature Genet. 3:219]. Rat livers have also been efficiently transduced *ex vivo* and then re-implanted [Shaked *et al.* (1994) Transplantation 57:1508]. Thus, the present invention contemplates *ex vivo* transfection followed by transplantation of the transfected cells or organ.

The replication defective recombinant adenoviruses are preferably employed; these viruses contain a deletion of the key immediate early genes E1a and E1b. To generate and propagate recombinant viruses, a packaging cell line such as 293 cells which supply the E1a and E2a proteins *in trans* is employed. Recombinant adenoviruses are created by making use of intracellular recombination between a much larger plasmid encoding most of the viral genome and a small plasmid containing the nucleotide sequence of interest flanked by

regions of homology with the viral integration site. Standard methods may be used to construct the recombinant adenoviruses [Graham and Prevec (1991) Meth. Mol. Biol. 7:109-128; Becker et al. (1994) In Protein Expression in Animal Cells, Roth et al. eds.]. Briefly, each plasmid is co-transfected together with pJM17 (Microbix Systems, Toronto) into sub-confluent monolayers of 293 cells (ATCC CRL 1573) using calcium phosphate precipitation and a glycerol shock. Initial recombinant viral stocks are titered on monolayers of 293 cells, and isolated single plaques are obtained and tested for expression of the polypeptide of interest using ELISA. Viral stocks are amplified and titered on 293 cells, and stored in aliquots at -70°C; if necessary, stocks are concentrated by centrifugation on density gradients. To infect tumor cells with recombinant adenoviruses, freshly isolated tumor cells are mixed with adenoviral stocks in a minimal volume. Titers of stocks are typically 10<sup>5</sup>- 10<sup>8</sup>/ml. Medium is replaced after several hours and the cells are followed for expression of the recombinant adenoviral-encoded polypeptide of interest (e.g., reporter genes).

A potential drawback of using an adenoviral delivery system is that the transduced cells may retain or express small quantities of adenoviral antigens on their surface. "Second generation" adenoviral vectors which contain deletions in the E2a gene are available and are associated with less inflammation in the recipient and a longer period of expression of the gene of interest [Engelhardt *et al.* (1994) Proc. Natl. Acad. Sci. USA 91:6196]. If necessary, nucleic acid sequences encoding polypeptides of interest are inserted into second generation adenoviral vectors.

Recently, adenoassociated virus (AAV) vectors and chimeric lentivirus vectors have also been shown promise in the expression of polypeptide sequences of interest.

#### 3. Targeted Cationic Liposomes

5

10

15

20

25

30

Cationic liposomes have proven to be a safe and effective means for inducing the transient expression of DNA in target cells [Ledley (1995) Human Gene Ther. 6:1129]. Clinical trials are underway using cationic liposomes to introduce the CFTR gene into the lungs of cystic fibrosis patients [Caplen et al. (1994) Gene Ther. 1:139 and Alton et al. (1993) Nature Genet. 5:135] or to introduce, by direct intra-tumor injection, the T cell costimulator B7-1 into malignant melanoma lesions in order to induce a cell-mediated immune response [Nabel et al. (1993) Proc. Natl. Acad. Sci. USA 90:11307].

Cationic liposomes (e.g., DOTAP/DOPE) and ligand-targeted cationic liposomes may be employed for the delivery of polypeptides of interest to tumor cells. Recently, in addition to cationic liposomes, neutral liposomes have also been reproted to also be useful in targeing ligands to cells. Ligand-targeted liposomes are made by covalently attaching ligands or antibodies to the surface of the cationic liposome. For example, when glioblastoma cells are to be targeted, transferrin is used as the ligand as glioblastoma cells express high levels of the transferrin receptor on their surface. When melanoma cells are to be targeted, internalizing receptors, monoclonal antibodies directed against melanoma-specific surface antigens (e.g., mAb HMSA5) may be employed as the ligand.

10

15

20

5

Plasmid DNA encoding polypeptides of interest is formed into a complex with preformed cationic liposomes using standard methodology or alternatively the DNA is encapsulated into the liposome interior. The DNA-containing liposomes are then used to transfer the DNA to tumor cells *in vivo* by direct intra-tumor injection or *in vitro* (using freshly explanted tumor cells) followed by return of the transduced cells to the recipient (e.g., a human patient or non-human animal).

# 4. Gene Transfer Using Biolistics

Biolistics (microballistics) is a method of delivering DNA into cells by projection of DNA-coated particles into cells or tissues. DNA is coated onto the surface of gold or tungsten microparticles (~1-3 µm diameter) and these particles are accelerated to high velocity and are impacted onto the target cells. The particles burst through the cell membrane and lodge within the target cell. The cell membrane quickly reseals and the passenger DNA elutes off of the particle and is expressed. The biolistic method has been used to transfect mammalian cells [Sanford *et al.* (1993) Methods Enzymol. 217:483].

25

A hand-held biolistic apparatus (BioRad) is used to transfer DNA into tumor cells or isolated tumor fragments. This device uses compressed helium to drive a disc-shaped macroprojectile which carries on its surface microparticles (1-5 µm) of gold which have been coated with purified plasmid DNA (coprecipitated with spermine) (Williams *et al.*, *supra*). This apparatus has been used to successfully transfect primary tissues.

30

Plasmid DNA encoding the polypeptides of interest may be coated onto the surface of gold microparticles according to the manufacturer's instructions (BioRad) and the biolistic apparatus is used to transfer the DNA into freshly explanted tumor cells or directly into

exposed tumors (e.g., metastatic nodules on the surface of the liver, melanoma lesions on the skin).

Regardless of the method of delivery of the expression vector into a cell, it is preferred, though not required, that the expression vector contain a selection marker (e.g., neo gene) to facilitate selection of transfected cells. Transfected cells are selected by growth in the presence of G418 (e.g., 200 μg/ml), followed by culture in growth medium containing reduced concentrations of G418 (e.g., 100 μg/ml) and growth to confluence. Expression of the polypeptides of interest is evaluated using, for example, immunoblot analysis or flow cytometry using monoclonal antibodies which are specific for the polypeptides of interest. It is preferred, though not necessary, that expression of the polypeptides of interest in the transfected tumor cells is both constitutive and stable. Constitutive expression refers to expression in the absence of a triggering event or condition, and can be achieved by the selection of a promoter which drives expression of the nucleic acid sequence encoding the polypeptides of interest. Examples of promoters which drive constitutive expression of a structural nucleic acid sequence which is operably linked to the promoter include the SRα promoter, CMV promoter, and HIV promoter.

Regardless of the type of expression vector used for delivery of the nucleic acid sequences of interest into a cell, the expression vector may be introduced to the cell by direct injection into tumor and/or preneoplastic tissue, systemic (e.g., intravenous) administration, aerosol administration (e.g., for delivery to the bronchial tree and other lung tissues), injection into breast ducts (e.g., for delivery to breast tissue), and topical administration (e.g., for delivery to cervical tissue).

### F. Reducing Expression Of Factor IX In An Animal

5

10

15

20

25

30

The regulatory sequences of the invention may also be used to reduce expression of a polypeptide sequence of interest which is encoded by a nucleic acid sequence whose transcription is under the regulatory control of the regulatory sequences provided herein. For example, the regulatory sequences of the invention may be used to reduce the rate of agerelated increase of EIX activity in an animal as a means of treating diseases (e.g., thrombosis, cardiovascular disease, etc.) which are associated with age-related increases in FIX activity. Since the inventors have discovered that the exemplary nucleic acid sequences AE5' and AE3' regulate stable and increased expression levels, respectively, of hFIX, the

increase in the level of hFIX activity over time may be reduced by inhibiting the function of AE3' which regulates increased expression of hFIX. This approach has the advantage that expression of hFIX remains under the control of AE5' thus providing hFIX activities which are stable over time and which continue to play an important role in normal blood coagulation processes.

5

10

15

20

25

**30** 

The function of AE3' in age-related expression of FIX may be inhibited by, for example, inhibiting the activity of the protein which specifically binds to AE3'. The protein(s) which bind to AE3' may be identified by using the AE3' (or the minimum portion of AE3' which has age-related regulatory activity) to screen protein libraries for specific binding to AE3' or its portion. Once the protein which binds to AE3' is identified, the function of this protein may be inhibited using antibodies which are specific for this protein. Antibodies which are specific for the protein which binds to AE3' are expected to disrupt the interaction between AE3' and this protein.

Antibodies (polyclonal and monoclonal) which are specific for the protein that binds to AE3' or portions thereof may be generated using methods known in the art. The term "antibody" refers to immunoglobulin evoked in animals by an immunogen (antigen). It is desired that the antibody demonstrates specificity to epitopes contained in the immunogen. The term "polyclonal antibody" refers to immunoglobulin produced from more than a single clone of plasma cells; in contrast "monoclonal antibody" refers to immunoglobulin produced from a single clone of plasma cells. The terms "specific binding," "specifically binding" and grammatical equivalents thereof when used in reference to the interaction of an antibody and an immunogen mean that the interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) on the immunogen; in other words the antibody is recognizing and binding to a specific immunogen structure rather than to immunogens in general. For example, if an antibody is specific for epitope "A", the presence of an immunogen containing epitope A (or free, unlabelled A) in a reaction containing labelled "A" and the antibody will reduce the amount of labelled A bound to the antibody.

Polyclonal and monoclonal antibodies which are specific to a desirable polypeptide, given the teachings herein, may readily be prepared by one of skill in the art. For example, monoclonal antibodies may be generated by immunizing an animal (e.g., mouse, rabbit, etc.) with a desired antigen and the spleen cells from the immunized animal are immortalized,

commonly by fusion with a myeloma cell. Immunization with antigen may be accomplished in the presence or absence of an adjuvant, e.g., Freund's adjuvant. Typically, for a mouse, 10 µg antigen in 50-200 µl adjuvant or aqueous solution is administered per mouse by subcutaneous, intraperitoneal or intra-muscular routes. Booster immunization may be given at intervals, e.g., 2-8 weeks. The final boost is given approximately 2-4 days prior to fusion and is generally given in aqueous form rather than in adjuvant.

5

10

15

20

25

30

Spleen cells from the immunized animals may be prepared by teasing the spleen through a sterile sieve into culture medium at room temperature, or by gently releasing the spleen cells into medium by pressure between the frosted ends of two sterile glass microscope slides. The cells are harvested by centrifugation (400 x g for 5 min.), washed and counted. Spleen cells are fused with myeloma cells to generate hybridoma cell lines. Several mouse myeloma cell lines which have been selected for sensitivity to hypoxanthineaminopterin-thymidine (HAT) are commercially available and may be grown in, for example, Dulbecco's modified Eagle's medium (DMEM) (Gibco BRL) containing 10-15% fetal calf serum. Fusion of myeloma cells and spleen cells may be accomplished using polyethylene glycol (PEG) or by electrofusion using protocols which are routine in the art. Fused cells are distributed into 96-well plates followed by selection of fused cells by culture for 1-2 weeks in 0.1 ml DMEM containing 10-15% fetal calf serum and HAT. The supernatants are screened for antibody production using methods well known in the art. Hybridoma clones from wells containing cells which produce antibody are obtained, e.g., by limiting dilution. Cloned hybridoma cells (4-5 x 106) are implanted intraperitoneally in recipient mice, preferably of a BALB/c genetic background. Sera and ascites fluids are collected from mice after 10-14 days.

The invention also contemplates humanized antibodies which may be generated using methods known in the art, such as those described in U.S. Patent Numbers 5,545,806; 5,569,825 and 5,625,126, the entire contents os which are incorporated by reference. Such methods include, for example, generation of transgenic non-human animals which contain human immunoglobulin chain genes and which are capable of expressing these genes to produce a repertoire of antibodies of various isotypes encoded by the human immunoglobulin genes.

5

10

15

20

25

30

PCT/US00/15728

Alternatively, the function of AE3' in age-related expression of FIX may be inhibited by, for example, inhibiting the activity of AE3' using antisense sequences which are directed to AE3'. The term "antisense" as used herein refers to a deoxyribonucleotide sequence whose sequence of deoxyribonucleotide residues is in reverse 5' to 3' orientation in relation to the sequence of deoxyribonucleotide residues in a strand of a DNA duplex. AE3' antisense sequences may be used to turn off genes under the expression regulation of AE3' by transfecting a cell or tissue with expression vectors which express high levels of a desired AE3' antisense oligomer (e.g., 15-20 nucleotides) or larger fragment. Such constructs can flood cells with antisense sequences which inhibit expression of FIX. Antisense sequences can be designed from various locations along the AE3' sequence. Animals (e.g., mice) treated with vectors expressing AE3' antisense sequences are monitored for changes in the age-related symptoms associated with FIX expression. The alleviation or treatment of one or more of these symptoms in animal by an antisense sequence suggests that the antisense sequence may be useful in the treatment and/or prevention of age-related FIX expression in humans.

#### **EXPERIMENTAL**

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof. Unless otherwise mentioned, all reference to nucleotide numbers with respect to the factor IX nucleotide sequence, refers to the nucleotide numbers of the hFIX gene sequence shown in Figure 8.

### **EXAMPLE 1**

# Construction Of A Series of Twelve Exemplary Human Factor IX (hFIX) Minigene Expression Vectors

To explore the molecular mechanisms underlying age-related regulation of Factor IX, a series of twelve hFIX minigene expression vectors were constructed. These vectors were first analyzed *in vitro* in HepG2 cells, a human hepatoma cell line (see Example 2, *infra*). Transgenic mice harboring the hFIX minigene vectors were generated and longitudinal analyses of hFIX expression for the entire life spans of founders and successive generations of transgenic mice were carried out (See Example 3, *infra*).

The twelve exemplary minigenes contained sequences derived entirely from the hFIX gene sequence, including (a) promoter sequences of various lengths spanning up to nucleotide (nt) -2231 in the 5' flanking region, (b) the coding region containing a first intron in which the first intron's middle portion is truncated. i.e., nt +1098 through nt +5882 of Figure 8, and (c) either the complete 3' UTR sequence or the 3' UTR sequence in which the middle portion was deleted. Figure 1 shows the structure of eleven out of the twelve human FIX minigene expression constructs. The name of each construct is shown at left. The structure is depicted with the promoter-containing regions (solid thick line on left) with the 5' terminal nucleotide number. Transcribed hFIX regions (open rectangles connected with thin lines representing the shortened first intron) are followed by 3' flanking sequence regions (solid thick line at right). Arrow: transcription start site; asterisk: translation stop codon; pA: polyadenylation; sl: potential stem-loop forming dinucleotide repeats.

5

10

15

20

25

30

Construction of hFIX minigene expression vectors was carried out using -416FIXm1 as the starting construct (Kurachi et al. (1995) J. Biol. Chem. 270:5276-5281). The nucleotide (nt) numbering system used in this study was based on the complete hFIX gene sequence previously reported (Yoshitake et al. (1985) Biochem. 24:3736-3750). Minigene -416FIXm1/1.4 was constructed from -416FIXm1 by inserting the middle portion of the 3' UTR (1.2 kb in size) which was generated by PCR using the following primer set with BamH I linkers: 5' primer, TAACAGGATCCGGCCTCTCACTAACTAATCAC (nt +31418 through +31438) (SEQ ID NO:14) and 3' primer, CAACTGGATCCAAGATTCAAGATAGAAGGAT (nt +32690 through +32671) (SEQ ID NO:15), and human genomic DNA as an amplification template. The PCR product was digested with BamH I, and the generated fragment was inserted into the 3' UTR BamH I site of -416FIXm1, thus producing -416FIXm1/1.4 which contained the entire 3' UTR. -416FIX m1/0.7 was constructed by inserting the PCR-amplified 700 bp fragment with BamH I linker, containing the 3' contiguous sequence to nt +32117. The primers used were, 5' primer: same as that for -416FIXm1/1.4, 3'primer: GGACAGGATCCCC CAAACTTTTCAGGCAC (nt +32117 through +32097) (SEQ ID NO:16). Minigenes -590FIXm1, -679FIXm1, -770FIXm1, -802FIXm1 and -2231FIXm1 were produced by replacing the 5' end 433 bp sequence of -416FIXm1 released by Sph I/Nhe I digestion with 607, 696, 787, 819 and 2248 bp fragments containing the 5' end hFIX region extended up to nt -590, -679, -770, -802 and -2231, respectively. These latter fragments were generated by

Sph I/Nhe I digestion of the PCR product obtained with 5' primers:

CAAGCATGCATCTAGTGTTAGTGGAAGAG (nt -590 through -571) (SEQ ID NO:17),

CAAGCATGCAAATATTAACTCAAAATGGA (nt -679 through -660) (SEQ ID NO:18),

CAAGCATGCTGTTGTTTTTGTTTTAAC (nt -770 through -752) (SEQ ID NO:19),

CAAGCATGCAGCCATTCAGTCGAGGAAGG (nt -802 through -783) (SEQ ID NO:20),

CAAGCATGCGATCCCTTCCTTATACCT (nt -2231 through -2214) (SEQ ID NO:21) with

Sph I linker and the common 3' primer TAAGCTTAACCTTTGCTAGCAGATTGT (nt +30 through +10) (SEQ ID NO:22) and human genomic DNA as the amplification template.

Minigene -802FIXm1/0.7 (whose structure is not shown in Figure 1) contains the 3' UTR region through nt 32,140, which is then connected to nt 32,690 through its downstream poly

-416FIXm1/AE5' depicts a construct with the AE5' region moved to the 3'-end position and shown as an open box at right. -416FIX m1/AE5' was constructed by inserting the Kpn I fragment generated by PCR (nt -802 through nt -417) into the -416FIXm1 vector (the Kpn I site is outside of the FIX gene, Figure1). The 5' and 3' primers used for PCR were CTTGGTACCAGCCATTCAGTCGAGGAAGG (nt -802 through -783) (SEQ ID NO:23) and CTTGGTACCATATGAATCCTTTCATAGAT, (nt -417 through -436) (SEQ ID NO:24) respectively. All constructs were sequenced through PCR amplified regions as well as fragment ligation sites to confirm the correct sequences and orientations.

(A) signal sequence that is common to each of the other eleven constructs.

20

5 .

10

15

### **EXAMPLE 2**

# Transient Expression of Eleven hFIX Minigene Expression Vectors In Vitro In Human Hepatoma HepG2 Cell Line

30

25

Transient *in vitro* expression activities of hFIX minigene constructs were assayed using HepG2 cells and hFIX specific enzyme linked immunosorbent assay (ELISA) as previously described (Kurachi *et al.* (1995) J. Biol. Chem. 270:5276-5281) with some modifications. Cell transfection was carried out by the calcium phosphate-DNA conjugate method or later, using FuGene 6 (Boehringer Manheim). The latter, improved transfection method consistently increased transfection efficiency to >20% (Kurachi *et al.* (1998) Biochemica 3:43-44), and all earlier assays were reexamined using FuGene 6. Four to five independent assays of factor IX activity were carried out and the averages were shown with

standard errors. With FuGene 6 transfection, the control minigene -416FIXm1 typically produced hFIX at a level of ~50 ng/10<sup>6</sup> cells/48 hr.

5

10

15

20

25

30

Figure 1 shows the relative *in vitro* transient expression activities of the human FIX minigene expression constructs (transient expression activity of minigene -802FIXm1/0.7 which is not shown in Figure 1 was 81.5% of the activity of minigene -416FIXm1). Transient expression activities relative to the activity of -416FIXm1 (~50 ng/10<sup>6</sup> cells/48 hour, and defined as 100% activity) are shown on the right side with standard deviations (from 4-5 independent assays). Activities were normalized to the size of minigenes used.

The relative transient expression activities shown in Figure 1 show that all constructs showed comparable high transient hFIX expression in HepG2 cells (~50 ng/10<sup>6</sup> cells/48 hours). However, all the constructs containing the complete 3' UTR, including a 102 base pair (bp) stretch of inverted AT, GT and GC dinucleotide repeats [Yoshitake et al. (1985) Biochem. 24:3736-3750], reproducibly showed expression activity levels which were 25-30% lower than corresponding minigenes without the repeat sequences. Dinucleotide repeats similar to those seen in the hFIX 3' UTR, which can form stable stem-loop (sl) structures in mRNA, have been implicated in controlling mRNA stability in mammals as well as yeast and plants, thus providing an important layer of protein biosynthesis regulation [Ross (1995) Microbiol. Rev. 59:423-450]. Together, these results suggest a similar negative regulatory activity for this structure of the hFIX gene in the HepG2 assay system on expression of the hFIX gene. As described below (e.g., Example 3), however, the 3' UTR structure of the hFIX gene containing the dinucleotide repeat region showed unexpected functions in vivo which are critical for advancing age-related regulation of the hFIX gene.

Another important and surprising finding with the HepG2 cell assay system is that expression by these hFIX minigenes (which contained sequences which are positioned upstream and downstream of the hFIX gene, and which are derived from the homologous hFIX gene instead of from heterologous reporter genes) does not show any down-regulation in the presence of the 5' upstream region (nt -802 up through nt -1900) [Salier *et al.* (1990) J. Biol. Chem. 265:7062-7068] (Figure 1). In contrast, when a CAT reporter gene was used, negative regulatory elements were identified in this region [Salier *et al.* (1990) J. Biol. Chem. 265:7062-7068].

#### **EXAMPLE 3**

### Generation And Analysis Of Transgenic Mice Harboring hFIX Minigene Expression Vectors

Transgenic animals were constructed using the expression plasmids described above in Example 2 according to standard methods [Hogan et al. (1994) in "Manipulating the Mouse Embryo, a Laboratory Manual" (Cold Spring Harbor Press, New York, 2nd Edition). All animal experiments were carried out in accordance with the institutional guidelines of the University of Michigan (OPRR No. A3114-01).

Briefly, Factor IX minigene expression plasmids were double-digested with Sph I/Kpn I and the factor IX minigene-containing fragments released were isolated by 0.8% agarose gel electrophoresis, followed by purification with SpinBind DNA extraction units (FMC). Fertilized eggs of C57B/6 X SJL mice were microinjected with the DNA (1-2 ng/egg), and implanted into foster mother animals (CD-1).

### A. Multiplex PCR Analysis

5

10

15

20

25

30 ·

Offspring produced were screened for founder animals with high transgene copy numbers (5 -10 copies per genome) using quantitative multiplex PCR analyses of tail tissue DNA samples. Two pairs of primers were used, one specific to the hFIX transgenes and the other specific to mouse  $\beta$ -globin gene (endogenous control); 5' primer:

CTGTGGGAACACACAGATTTTGG (nt +6172 through +6195) (SEQ ID NO:25) and 3' primer: GGAATAATTCGAATCACAT (nt +30885 through +30867) (SEQ ID NO:26), and 5' primer: CCAATCTGCTCACACAGGAT (nt +2590 through +2609) (SEQ ID NO:27) and 3' primer: CCTTGAGGCTGTCCAAGTGA (nt +3083 through +3064) (SEQ ID NO:28), respectively. These primers were designed to amplify a unique 966 bp fragment from the hFIX transgenes and a 494 bp fragment from the mouse β-globin gene, respectively. PCR was initiated with 3 min incubation at 94° C, followed by 25 cycles of 94° C for 30 sec, 65° C annealing for 1 min and 72° C extension for 2 min.

Founders were back-crossed with non-transgenic mice (C57B/6 X SJL) to generate F1 progeny animals. Homozygous F2 animals were generated by crossing among heterozygous F1 littermates and the following generations were similarly generated. Zygosity status of animals was determined by quantitative multiplex PCR analysis as described above.

Minimally, three founder lines for each minigene construct were subjected to longitudinal analysis for their entire life spans up to two years.

5

10

15

20

25

30

Figure 3B shows the results of quantitative multiplex PCR analysis to determine the relative transgene levels in tail and liver tissues. Genomic DNA was extracted from snipped tail tissue of a transgenic -416FIXm1 animal (PA112) at 3 weeks and at 19 months of age. Liver DNA was extracted from the same animal (PA112) sacrificed at 19 months of age and a -416FIXm1 animal (PA412) sacrificed at 1 month of age. Positions of hFIX specific fragment (966 bp) and mouse β-globin specific fragment (494 bp, internal copy number control) are shown on the right. Lane 1: kb size ladder; lane 2: fragment size control amplified from -416FIXm1 plasmid; lane 3: non-transgenic mouse tail DNA as template; lane 4: tail DNA of PA112 at 3 weeks of age; lane 5: tail DNA of PA112 at 19 months of age; lane 6: liver DNA of PA412 at 1 month of age; lane 7: liver DNA of PA112 at 19 month of age. The relative transgene copy numbers for the 1 month-old versus the 19 month-old animals, normalized to the endogenous mouse β-globin gene, were 1.0-1.1 for both tail as well as liver genomic DNA preparations, showing no sign of loss of the hFIX transgene in the genome with age (Multi Analyst program from BioRad used for quantitation and calculation of ratios).

### B. Immunoassay of hFIX Levels In Transgenic Mice

Circulatory hFIX levels were monitored during longitudinal analyses of transgenic mice from the representative founder lines carrying various hFIX minigene transgenes. At various ages, starting at one month of age, transgenic mice were individually subjected to blood sample collection (aliquot of ~100 µl) via tail-tip snipping, and the obtained serum was routinely used to quantify hFIX levels in the circulation using duplicated hFIX-specific ELISA for each age point. Pooled human plasma (George King Bio-Medical) was used to prepare a hFIX standard curve for each assay. In order to minimize experimental fluctuations from assay to assay in the longitudinal analysis, overlapped serum samples from the previous assay group were included in each assay. To ensure reproducibility, three to six independent founder lines were generated for each minigene construct, and animals from at least three representative lines were subjected to longitudinal analyses for their entire life spans. The duplicated ELISA values varied less than 11% from the averages. The results are shown in Figures 2 and 4.

In all panels in Figures 2 and 4, labeling of animals is based on the tag numbers plus additional information. The first letters of the label F or P represent founder or progeny, respectively. Information on progeny generation (F1 or F2) and sex are in parenthesis (m: male; f: female), followed by status (+: alive in good health; d: died; s: sacrificed for various examinations; mo: age of death or sacrifice). To avoid overcrowding of the panels, the results from representative animals are shown for each minigene construct. Importantly, age-regulation patterns were remarkably similar among all animals for each specific construct and different founder lines. Panels A-E of Figure 2 show representative founder line animals with -416FIXm1 (A); -416FIXm1/1.4 (B), -590 FIXm1 (C), -679FIXm1 (D) and -770FIXm1 (E). Panels A-D of Figure 4 show representative founder line animals with -802FIXm1, -802FIXm1/1.4, -2231 FIXm1 and -2231FIXm1/1.4, respectively. Panel E shows representative founder line animals with -416FIXm1/AE5'.

5

10

15

20

25

30

Figure 2 shows that at one month of age, the mice carrying the -416FIXm1 minigene produced hFIX at varying levels, from as high as that of natural hFIX gene expression (~4 μg/ml) to much lower levels (~50 ng/ml) (Figure 2A). Such variations are primarily due to the transgene positional effects in the genome. Circulatory hFIX levels of animals from the representative founder lines carrying the minigene, however, declined drastically through puberty and during the subsequent two to three month period to much lower levels, which then remained stable for the remaining life span. This rapid age-dependent characteristic decline in the circulatory hFIX level was observed in all animals analyzed (n=69), regardless of founder line, differences in initial hFIX level at pre-pubertal age (one month) due to transgene positional effects, generation (founders and F1 or F2 progeny), sex, or zygosity status (homozygous / heterozygous) of the transgenes.

### C. Northern Blot Analysis of hFIX mRNA in Transgenic Mice

Northern blot analyses of the liver RNA samples from animals (15 µg per lane) were carried out as previously described [Kurachi et al. (1995) supra] using the <sup>32</sup>P-labeled Ssp I/BamH I fragment (the 3' half of the hFIX coding region of the cDNA) as a probe, and employing stringent washing conditions. Under these conditions, the probe preferentially hybridized strongly with hFIX minigene mRNA bands (~1.7 kb) with little cross-hybridization with the mouse FIX mRNA bands (3.2 kb and 2.2 kb) [Yao et al. (1994) Gene Therapy 1:99-107]. To confirm the presence of equivalent amounts of RNA in each lane, the filters

previously hybridized with hFIX probe were stripped of probe and re-probed with the RNR18 cDNA (ribosomal RNA 18S). After completion of longitudinal analyses of animals from key founder lines for their entire life spans, the representative lines were subjected to embryo-freezing for preservation.

5

10

15

20

25

30

The results of Northern analysis of human FIX mRNA and transgene DNA levels in the livers of animals carrying -416FIXm1 are shown in Figure 3A. hFIX mRNA levels in the liver of young (PA412: F1/f, 1 month of age) and old (PA112; F1/f, 19 months of age) transgenic animals were analyzed by Northern blot analysis of total liver RNA. PA412 and PA112 animals were from the same litter produced by the founder FA661, and expressed 1252 and 1675 ng/ml circulatory hFIX at one month of age, respectively. PA112 was expressing 63.8 ng/ml serum hFIX at the time of sacrifice. Lane 1: non-transgenic mouse liver RNA; lane 2: transgenic PA412 liver RNA; lane 3: transgenic PA112 liver RNA. FIX and 18S on the left or right sides indicate the band position of hFIXm1 mRNA (~1.7 kb) and RNR18 (1.9 kb, ribosomal RNA), respectively.

Figure 3A shows that the decline in blood hFIX level observed in Figure 2 was correlated with a similar decline in the steady-state liver hFIX mRNA, which was not due to a loss of the hFIX transgene with age (Figure 3B). This was further supported by the fact that when 4-5 month old mice with much decreased hFIX levels had progeny, their pups depicted pre-pubertal high hFIX expression levels equivalent to those of their parents at the same time point (one-month of age).

Minigene vector -416FIXm1/1.4 is identical to -416FIXm1 except that -416FIXm1/1.4 contains the complete 3'UTR, including the dinucleotide repeat structure (102 bp in length) in its middle region [Yoshitake *et al.* (1985) Biochem. 24:3736-3750] (Figure 1). Transgenic mice with -416FIXm1/1.4 (n=48) (Figure 2B) showed pre-pubertal high and subsequent age-dependent decline in hFIX levels similar to those of -416FIXm1 (Figure 2A), although the decline was less steep and expression levels were stabilized at significantly higher levels than those observed for -416FIXm1 (Figure 2B).

These results indicate that, while the 102-bp sequence containing the dinucleotide repeat structure of hFIX 3' UTR reduces the age-related decline in expression of hFIX, the presence of the complete 3' UTR containing the extensive dinucleotide repeat structure nonetheless does not completely rescue hFIX expression from the age-decline observed in all

of these animals, regardless of founder line, initial pre-pubertal hFIX level, generation, sex, or zygosity status of the transgenes.

All animals carrying minigenes -590FIXm1 and -679FIXm1 (a total of 25 and 26 animals subjected to longitudinal analysis, respectively) also showed an age-associated rapid decline in hFIX expression similar to that seen in animals carrying -416FIXm1 (Figure 2, C and D). Furthermore, hFIX expression levels in three independent founder animals generated to date carrying -770 FIXm1 also rapidly decreased over the puberty period in a similar pattern as the above minigenes (Figure 2E). These observations indicated that minigenes with the promoter region up to nt -770 contain the basic structural elements necessary for hFIX expression, but lack a structural element(s) which functions in age-associated stability of hFIX gene expression.

5

10

15

20

25

30

In contrast, striking and unexpected differences in hFIX expression patterns were observed with animals carrying the minigene -802FIXm1 (Figure 4A) as comp[ared to those carrying the minigene -416FIXm1 (Figure 2A). -802FIXm1 is composed of a vector frame identical to -416FIXm1, except that the 5' end flanking sequence included was extended to nt -802 (Figure 1).

All animals with -802FIXm1, -2231FIXm1 and -416FIXm1/AE5' (panels A, C, E) exhibited stable expression throughout their life spans. Animals with -802FIXm1/1.4 and -2231FIXm1/1.4 (Figure 4 B, D) exhibited age-associated increases in hFIX expression levels. All animals maintained or increased stable circulatory hFIX levels regardless of founder line, initial expression levels at one month of age, sex, generation or zygosity status. Mice which died at much younger ages than their normal life expectancies are marked with d. The above results show that all animals from three independent founder lines obtained with -802FIXm1 (Figure 4A) showed characteristic differences in hFIX expression pattern from animals with -416FIXm1 (Figure 2A) and -416FIXm1/1.4 (Figure 2B).

The -802FIXm1 transgenic animals (n = 62) subjected to longitudinal analysis invariably showed age-stable plasma hFIX levels for their entire life spans, mostly up to 20-24 months of age. Age-stable circulatory hFIX levels were correlated with age-stable mRNA levels (Figure 5). These observations with -802FIXm1 were further supported by age-stable hFIX expression by mice carrying -2231FIXm1 (Figure 4C). Together, these results suggest that the structural element which is responsible for age-stable expression of the hFIX gene resides in the small region spanning nt -770 through -802. We designated

this small region "age-regulatory element in the 5' end" (AE5'). This region contains a transcription factor PEA-3 nucleotide sequence (GAGGAAG: nt -784 through -790), which completely matches the consensus motif (C/G)AGGA(A/T)G [Martin et al. (1988) Proc. Natl. Acad. Sci. 85:5839-5843; Xin et al. (1992) Genes & Develop. 6:481-496; Chotteau-Lelievre et al. (1997) Oncogene 15:937-952; Gutman and Wasylyk (1990) EMBO J. 9:2241-2246]. The function of AE5' nucleotide sequence is position-independent as shown by age-stable hFIX expression by animals containing -416FIXm1/AE5', in which AE5' was moved to the 3' end outside of the hFIX minigene (Figure 4E).

5

10

15

20

25

30

Since transgenes of -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1 differ from the minigenes -802FIXm1 and -2231FIXm1 only by their promoters, the hFIX mRNA produced from all of these minigenes (an intron spliced form of FIXm1 RNA) was expected to produce identical hFIX protein. Thus, it was hypothesized that the age-dependent decline in the circulatory hFIX level observed in animals with -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1, but not with -802FIXm1 and -2231 FIXm1, must be due to an age-dependent decline in the transcriptional activity of the transgenes. This agrees with the facts that no significant changes with age in hFIX mRNA levels in the liver were observed for animals carrying -802FIXm1 (Figure 5, lanes 2 and 3), while advancing age-dependent declines in steady-state mRNA level were observed for -416FIXm1 (Figure 3A, lanes 2 and 3).

To further determine whether the age-dependent decline in the circulatory hFIX levels was due to an age-dependent decline in transcriptional activity of the transgenes, the effects of age on hFIX clearance from the circulation were tested as follows. Aliquots of plasmaderived hFIX preparation (5 μg/0.1 ml of PBS) were injected via tail vein into normal animals at 2, 9-10 and 19-23 months of age (n=3 per age group), which have the same genetic background as the transgenic mice (C57B/6 X SJL). The hFIX level in circulation was monitored by ELISA of collected blood samples (~50 μl aliquot) at 10 min, 2, 6, 12, 18, 24, 30, 36 and 48 hrs after protein injection. As expected, all animals of different age groups showed a typical bi-phasic clearance kinetics (two compartment distribution and clearance) with an initial rapid clearance phase (α-phase), followed by a slower clearance phase (β-phase). The results are shown in Table 1.

Table 1
Clearance Time of Human Factor IX in Mice

Age (months)	Clearance Time (T <sub>1/2</sub> of Human Factor IX)				
2	16.8 ± 0.21				
9-10	17.4 ± 0.55				
19-23	16.9 ± 0.35				

5

10

15

20

25

30

As shown in Table 1, very similar half clearance times were observed for all age groups tested. This agreed with our previous results (17.8 hours) for hFIX clearance in a different strain, BALB/c mice (2 months of age) (Yao et al. (1994) supra].

Furthermore, the results in Table 1 demonstrate that the hFIX turnover time from the circulation does not change significantly *in vivo* with increasing age, from youth (2 months), to middle age (9-10 months) to old age (19-23 months). These results further confirm that the age-dependent decline in the circulatory hFIX levels was due to an age-dependent decline in transcriptional activity of the transgenes.

It is important to note that in the *in vitro* HepG2 cell assay system, the presence or absence of AE5' in the minigenes did not make any significant difference in hFIX expression from the hFIX minigenes (Figure 1, and Example 2, *supra*). In contrast, as mentioned above in this Example, the presence or absence *in vivo* of AE5' makes a dramatic age-dependent difference in hFIX gene expression. This further demonstrates that *in vivo* longitudinal analysis is important for studying age-regulation of a gene.

Unlike the animals with -802FIXm1, mice with -802FIXm1/1.4 (which contains the complete 3'UTR) showed an advancing age-associated increase in the hFIX level in the circulation (n=48) (Figure 4 A and B). Thus, to determine whether this unexpected age-dependent increase in the circulatory hFIX level was directly correlated with an increased level of liver hFIX mRNA, Northern blot analyses of transgenic mice carrying -802FIXm1 and -802FIXm1/1.4 were conducted. The hFIX mRNA levels in the liver of 1-month (young) or 15-month (aged) mice carrying -802FIXm1 (mouse P327 or P552, respectively) and -802FIXm1/1.4 (mouse P32 and P697, respectively) are shown in Figure 5. These

animals were from the same litter produced by the founder F17549 for -802FIX m1 and F229 for -802FIXm1/1.4 (Figure 4 A and B). At the time of sacrifice, P552 and P697 were expressing 2200 and 1658 ng/ml of hFIX, respectively. The total liver RNA (15 µg from each animal was used for the Northern blot analysis performed as described in Figure 3A. Upper panel: probed with the Ssp I/BamH I fragment of hFIX cDNA; lower panel: rehybridized with RNR18 (ribosomal RNA) probe. Lane 1: non-transgenic mouse liver; lane 2: transgenic P327 liver RNA; lane 3: transgenic P552 liver RNA; lane 4: transgenic P32 liver RNA; lane 5: transgenic P697 liver RNA. PhosphorImager (Molecular Dynamics) was used for quantitation of mRNA levels (counts) and ratios of young versus old were calculated. Young and old animals carrying -802FIXm1 showed no significant differences in the mRNA level (the ratio of old over young: 0.92). In contrast, -802FIXm1/1.4 animals showed a substantial elevation in the mRNA level with older age (the ratio of old over young: 1.54).

5

10

15

20

25

30

These results (Figure 5, lanes 4 and 5) indicated the presence of another important age-regulatory nucleotide sequence, designated AE3', which is located approximately in the middle of the 3' UTR where an extensive stretch of dinucleotide repeating structures were contained. In the presence of AE5', AE3' clearly confers a crucial age-associated increase in hFIX expression. This conclusion was further supported by results obtained with -2231FIXm1 /1.4 (n=42) (Figure 4D). The unique concerted function conferred by the combination of AE5' and AE3' was again independent of founder line, initial expression levels at one month of age, sex, generation, or zygosity status of animals.

Interestingly, animals with sustained high hFIX levels in the circulation (approximately 1,500 ng/ml or higher) tended to die at a much earlier age than the expected life span (~2 years) (Figure 4 A, B, D). This happened to both males and females, but appears to be more frequent in males. Without limiting the invention to any particular mechanism, it is believed that since these transgenic mice have hFIX in addition to their own mFIX, they may be at an increased risk of lethal thrombosis compared to wild type mice which do not express the transgenes.

The above-described characterization of transgenic mice harboring hFIX transgenes demonstrates that (a) while the presence of AE5' in vitro in HepG2 cells did not affect hFIX gene expression, the presence of AE5' in vivo resulted in a dramatic age-dependent increased stability in hFIX gene expression, (b) the age-dependent decline in the circulatory hFIX level

observed in animals with -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1 is directly correlated with the decrease in the steady-state mRNA level, which the inventors believe to be due to an age-dependent decline in the transcriptional activity of the transgenes, and (c) animals carrying -802FIXm1/1.4 shwoed a substantial elevation in the liver mRNA levels of hFIX with older age.

#### EXAMPLE 4

### Footprint And Gel Electrophoretic Mobility Shift Analysis Of The Region From Nucleotides -665 To -805 of Human Factor IX

In order to make a preliminary determination of the region within AE5' which is involved in the function of AE5', footprint analysis and gel electrophoretic mobility shift assays were performed as follows.

### A. Footprint Analysis

5

10

15

20

25

30

For footprint analysis of the region spanning from nt -665 through nt -805, the fragments used were amplified by PCR with the <sup>32</sup>P-labeled 5' primer ATGGTTAACTGACTTACGAA (nt -833 through -814) (SEQ ID NO:29) and 3' unlabeled primer GCTCCATTTTGAGTTAATATTTGTGT (nt -657 through -682) (SEQ ID NO:30). The nuclear extracts (NEs) from HepG2 human hepatoma cells and livers of young (1 month of age) and old (19 months of age) mice were prepared as previously reported [Kurachi et al. (1986) Biochemistry 33:1580-1591]. Various amounts of NEs (0, 100 and 150 μg) were incubated with the labeled fragments (30,000 CPM) for 1 hour on ice and subjected to DNase 1 digestion (0.5 unit) for 2 min at room temperature. The samples tested included those without NEs, with 100 μg and 150 μg of HepG2 cell NEs, with 100 μg and 150 μg NEs from young mice. Major and minor footprints and apparent DNase hypersensitivity sites were observed.

Footprint analysis of the region nt -665 through -805 with aged mouse liver nuclear extracts showed a major footprint (nt -784 through -802), a minor foot print (nt -721 through -728) and an interesting DNase hypersensitive region (nt -670 through -714). With nuclear extracts from one month-old animals or HepG2 cells, no such clear footprints were observed.

### B. Gel Electrophoretic Mobility Shift Assay

5

10

15

20

25

30

Gel electrophoretic mobility shift assay using mouse liver nuclear extracts from three different age groups was used. Nuclear extracts were prepared from 1, 5 or 19 month-old mouse livers (as described *supra*). Double stranded oligonucleotides containing a PEA-3 nucleotide sequence spanning from nt -797 to -776 of the hFIX gene (TTCAGTCGAGGAAGGATAGGGT) (SEQ ID NO:31) were <sup>32</sup>P-labeled at the 5' end to a specific activity of 1.9x10<sup>9</sup> cpm. Aliquots of the radio-labeled oligonucleotide (20,000 cpm) were incubated with 10 µg of NEs in the presence of 1 µg of poly dI-dC in DNA binding buffer for 20 min at room temperature and subjected to polyacrylamide gel electrophoresis (Kurachi et al. (1986) *supra*). In Figure 6A, Lane 1: without NEs; lane 2: with NEs of 1 month-old mice; lane 3: with NEs of 5 month-old mice; land 4: with NEs of 19 month-old mice; lane 5: with mouse brain NEs (positive control for PEA-3, showing a slightly higher size of shifted band).

Figure 6B shows the results of the competition assay for <sup>32</sup>P-labeled double stranded oligonucleotides containing the PEA-3 nucleotide sequence. A 100-fold excess unlabeled oligonucleotide described in the preceding paragraph or mutant oligonucleotide [TTCAGTCGGTTGGTGATAGGGT (SEQ ID NO:32) with mutated sequences underlined] was incubated with 10 µg of 19 month-old mouse liver NEs for 5 min followed by addition of <sup>32</sup>P-labeled oligonucleotides as described *supra*. Lane 1: without NEs; lane 2: with NEs; lane 3: with NEs and wildtype competitor; lane 4: with NEs and mutant competitor.

In agreement with the above results of footprinting, gel electrophoretic mobility-shift (bandshift) assays showed an increase in protein binding with the nuclear extracts of aged mice (19 months of age) (Figure 6A). Bandshifts were competitively reduced with excess amounts of oligonucleotides harboring the PEA-3 motif, but not with oligonucleotides harboring a mutant PEA-3 motif sequence (Figure 6B), thus confirming the presence of the PEA-3 motif in AE5'. This is the first time that the PEA-3 protein, which is a member of the Ets family of transcription factors and which has been shown to bind to nucleotide sequences [SEQ ID NO:40; SEQ ID NO:48; and SEQ ID NO:84] that are homologous to the PEA-3 nucleotide sequence within the AE5' region [Karim et al. (1990) Genes & Develop. 4:1451-1453; Nelsen et al. (1993) Science 261:82-86; Fisher et al. (1991) Oncogene 6:2249-2254], has been implicated in such a unique role in age-stable regulation of a gene.

5

10

15

20

25

30

Without limiting the invention to any particular mechanism, the PEA-3 nucleotide sequence in the hFIX gene appears to have been generated through evolutionary drift of a L1 sequence originally recruited presumably via its retrotransposition into the 5' specific location. Modern retrotransposable L1 [Kazazian et al. (1988) Nature 332:164-166; Dombroski et al. (1993) Proc. Natl. Acad. Sci. USA 90:6513-6517; Minakami et al. (1992) Nucl. Acids Res. 20:3139-3145; Dombroskiet et al. (1994) Mol. Cell. Biol. 14:4485-44921 does not have the corresponding PEA-3 nucleotide sequence. The PEA-3 nucleotide sequence of AE5' nucleotide sequence resides within the L1-derived sequence retaining a 63-70% similarity with the ORF2 corresponding region of the modern retrotransposable L1 in the 5' to 3' orientation. Interestingly, the murine FIX gene also has the L1-derived nucleotide sequence in its 5' end region in an almost identical position as in the hFIX gene. and has multiple PEA-3 consensus nucleotide elements [Kawarura et al.in Organization of L1 Sequence in the 5' Flanking Region of Factor IX Gene [in preparation]. Age-regulation of the murine FIX gene is indeed very similar to that of the hFIX gene [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi et al. (1996) Thromb. Haemost. 76:965-969], thus providing further insights into the evolutionary origin of the molecular mechanisms underlying age-associated regulation of the FIX gene.

### **EXAMPLE 5**

### Liver-Specific Expression Of The Exemplary hFIX Gene Under Control Of The hFIX Promoter

Expression of the natural FIX gene is virtually restricted to the liver [Salier et al. (1990) J. Biol. Chem. 265:7062-7068]. In order to determine whether any of the upstream and/or downstream sequences in the hFIX minigenes directed liver-specific expression of the hFIX transgene, Northern blot analysis was carried out as described supra (Example 3) in transgenic mice carrying -416FIXm1 and -802 FIXm1 expression vectors. Animals expressing hFIX at high level (PA412 and P580 carrying -416FIXm1 and -802FIXm1, respectively) were sacrificed at one month of age and total RNA was extracted from liver, lung, intestine, muscle, kidney, brain and heart and from untransfected HepG2 cells (negative control). The results in transgenic mice carrying -416FIXm1 and -802 FIXm1 are shown in Figure 7A and B, respectively.

In Figure 7, the positions of hFIX mRNA, RNR18 (control for RNA loading in wells), and ribosomal 28S and 18S RNA bands are shown on the left and right sides, respectively. Animals with -416FIXm/1.4 and -679FIXm1 showed tissue specific expression patterns similar to that of -416FIXm1 (A) (data not shown). Interestingly, liver expression of hFIX observed for minigenes lacking the region containing AE5' (except -770FIXm1, which remains to be tested as sufficient progeny animals become available) was high, but not as robust, as that seen with the natural gene. In addition, these minigenes expressed not only in the liver, but also in other tissues, such as kidney, lung and muscle, at various levels as high as ~20% of the liver level (Figure 7A). In clear contrast, animals with -802FIXm1 showed substantially liver-specific hFIX expression similar to that for the natural FIX gene (Figure 7B). These results suggest that the AE5' region controls liver specific expression of hFIX.

An apolipoprotein(a) transcription cotnrol region (ACR) which contains an ETS family target sequence 5'-CCCGGAAG-3' (SEQ ID NO:48) has been shown to exhibit enable enable activity in vitro in liver-derived HepG2 cells. However, the ACR does not appear to be liver-specific [Yang et al. (1998), supra].

### EXAMPLE 6

### Expression Of PEA-3 Protein In HepG2 Liver Cells

20

25

30

5

10

15

Expression of the transgene FIX was observed *in vivo*, but not *in vitro* in HepG2 cells, when expression vectors containing AE5' were used (See, Examples 2 and 3, *supra*). This observation, together with the absence of a footprint in HepG2 cell NEs (See, Example 4, *supra*) suggested to the inventors that HepG2 cells' lack of expression of the FIX transgene may be a result of the cells' expression of low levels of the PEA-3 protein (and/or the PEA-3 related protein) which binds to homologs of the invention's PEA-3 nucleotide sequence. The complete human PEA-3 cDNA has not yet been cloned (the human PEA-3 cDNA sequence of GenBank accession number U18018 lacks 8 amino acids at the N-terminal region when compared to the mouse PEA-3 cDNA sequence). In order to determine the role of the PEA-3 nucleotide sequence in gene expression *in vitro*, HepG2 cells which overexpress mouse PEA-3 protein were constructed as follows.

Expression constructs containing the mouse PEA-3 cDNA sequence (GenBank Accession Number X63190; Figure 9) were constructed as follows. Using the reported mouse PEA-3 cDNA sequence three sets of PCR primers were synthesized such that the entire coding region and parts of the flanking sequences would be amplified. Reverse transcription PCR (RT-PCR) was carried out, and the amplified mouse PEA-3 cDNA sequence was inserted into an expression vector under the control of the SV40 promoter, which does not interfere with the factor IX promoter (data not shown).

The PEA-3 expression vector is used to transfect HepG2 cells using the FuGene 6 (Boehringer Manheim) since this method was shown to improve transfection efficiency (See, Example 2, supra). Transfected HepG2 cells are screened for expression of PEA-3 by Northern blot analysis and/or Western blot analysis using commercially available antibody. Transfected HepG2 cell lines which stably express PEA-3 protein are selected for further use, e.g., to analyze the underlying mechanism of PEA-3 action.

15 EXAMPLE 7

5

10

20

25

30

### In vitro And In Vivo Expression Of Exemplary Human Protein C Minigene Expression Vectors Containing AE5' And AE3'

Protein C is a factor which plays a critical role in the anti-blood coagulation mechanism. Unlike factor IX, whose level in the circulation substantially increases with advancing age, protein C levels in the circulation do not increase with advancing age, but rather show a slight decrease over time. This decrease in circulating protein C levels is believed by the inventors to be the result of regulation at the gene transcription level. For this reason, the protein C gene provides an interesting exemplary gene for demonstrating the universality of the AE5' and AE3' function in gene expression both *in vitro* and *in vivo* as follows. In this Example, bases are numbered relative to the major transcription start site (+1) as previously described [Miao et al. (1996) J. Biol. Chem. 16:9587-9594].

#### A. Construction of Human Protein C Minigene Expression Vectors

The human protein C genomic sequence has been previously reported (GenBank accession number M11228; Figure 12B]. Using this sequence, three protein C minigene expression vectors were prepared. The first human protein C minigene vector (-1426PCm1)

contained the human protein C promoter region of the protein C gene (GenBank accession number M11228; Figure 12 B) ligated to the human protein C cDNA (GenBank accession number X02750; Figure 12 A) which contains the first entire intron and poly-A sequence. The second human protein C minigene (AE5'/-1426PCm1) was the same as the first vector except that it additionally contained the nucleotide sequence AE5' at the 5' end of the human protein C cDNA. The third human protein C minigene (AE5'/-1426PCm1/AE3') was the same as the first vector vector except that it additionally contained the nucleotide sequences AE5' and AE3' at the 5' and 3' ends, respectively, of the human protein C cDNA.

10

15

20

25

30

**5** 

### B. Transient Expression Of human Protein C in vitro In HepG2 Cells

Each of the protein C minigene expression vectors was transfected into HepG2 cells using the FuGene 6 (Boehringer Mannheim). Four to five independent assays of human protein C activity were carried out as previously described [Turkey et al (1999) Throm. Haemost. 81:727-732]. HepG2 cells transfected with the -1426PCm1 vector showed *in vitro* transient protein C activities which were comperable to the activities shown by HepG2 cells transfected with the AE5'/-1426PCm1 vector (*i.e.*, (60-70 ng/10E6 cells/24 hrs).

# C. Generation Of Transgenic Mice Harboring The protein C Minigene Expression Vectors

In order to determine whether AE5' in combination with AE3' is capable of increasing human protein C expression with advancing age, as observed for factor IX expression (Example 3, supra), transgenic mice which harbor the protein C minigene expression vectors are generated according to standard methods [Hogan et al. (1994), supra]. Briefly, protein C minigene vector plasmids are injected into fertilized eggs of C57B/6 X SJL mice and implanted into foster mother animals (CD-1). Offspring produced are screened for founder animals with high transgene copy numbers (5 -10 copies per genome) using quantitative multiplex PCR analyses of tail tissue DNA samples using two pairs of primers which are designed to amplify a unique fragment from the protein C transgenes and a 494 bp fragment from the mouse β-globin gene, respectively.

Founders are back-crossed with non-transgenic mice (C57B/6 X SJL) to generate F1 progeny animals. Homozygous F2 animals are generated by crossing among heterozygous

F1 littermates and the following generations are similarly generated. Zygosity status of animals is determined by quantitative multiplex PCR analysis as described above. Founder lines for each minigene construct are subjected to longitudinal analysis for their entire life spans up to two years.

Circulatory human protein C levels are monitored during longitudinal analyses of transgenic mice from the representative founder lines carrying the protein C minigene transgenes. Age-regulation patterns of circulatory human protein C levels are compared among all animals for each specific construct, different founder lines, different initial human protein C level at pre-pubertal age (one month) due to transgene positional effects, generation (founders and F1 or F2 progeny), sex, and zygosity (homozygous / heterozygous) status of the transgenes.

Northern blot analyses of the liver RNA samples from animals is carried out using stringent washing conditions to determine whether any changes in circulatory human protein C levels are correlated with similar changes in the steady-state liver human protein C mRNA, rather than with loss of the human protein C transgene with age or with changes in human protein C turnover time. Observation of transgenic animals which contain the human protein C sequence as well as AE5' and AE3' and which increase stable circulatory human protein C levels with increasing animal age, as compared to the levels in transgenic animals which express the human protein C sequence in the absence of AE5' and AE3', demonstrates that the combination of AE5' and AE3' functions in age-stable expression of the exemplary human protein C gene. This observation will confirm that these results may be achieved for genes other than the exempary hFIX and protein C genes.

#### **EXAMPLE 8**

25

30

5

10

15

20

# In Vivo Expression Of Exemplary Expression Vectors Containing The Cytomegalovirus (CMV) Promoter And The AE5' And AE3'

This Example is carried out to demonstrate the universality of the age-related gene expression regulatory function of AE5' and AE3' with viral promoters. The CMV promoter is currently used in several gene therapy constructs but its activity decreases with time in the Liver. Furthermore, the activity of the CMV promoter in the liver of transgenic mice is known to be lower than the activity in other tissues, such as muscle. Thus, this Example

investigates whether the combination of AE5' and AE3' halts or reverses the decline in the activity of the CMVpromoter in the liver.

The above-described -416FIXm1 expression vector (Example 1, and Figure 1) is used to construct a control vector to determine the effect of AE5' and AE3' on liver and circulatory levels of expression of human factor IX in transgenic animals. The control vector in which expression of the hFIX gene is under the control of the CMV promoter in the absence of both AE5' and AE3' is constructed by replacing the human factor IX promoter sequence with the CMV promoter sequence (National Vector Core for Non-Viral Vectors at the University of Michigan) (the CMV promoter is also located between positions +1 to +596 in vector plasmid pCR3 from Invitrogen). The resultant expression vector in which the human factor IX gene is under the control of the CMV promoter is transfected into HepG2 cells. Transfected cells are expected to show human factor IX activity.

5

10

15

20

25

30

The -802FIXm1/1.4 vector of Figure 1 is used to construct a test vector in which the human factor IX promoter sequence of -802FIXm1/1.4 vector (which contains both AE5' and AE3') is replaced with the CMV promoter sequence.

In order to determine whether the combination of AE5' and AE3' is capable of increasing human factor IX expression with advancing age under the control of the CMV promoter, as observed for factor IX expression under the control of the factor IX promoter (Example 3, supra), transgenic mice which harbor either the control vector or the test vector are generated according to standard methods as described supra (Examples 3 and 7). The mRNA levels of human factor IX in the blood, liver and other tissues are monitored during longitudinal analyses of the transgenic mice. Age-regulation patterns of human factor IX mRNA levels in the different tissues are compared among all animals as described supra for each specific construct, different founder lines, different initial human factor IX levels at prepubertal age due to transgene positional effects, generation, sex, and zygosity status of the transgenes.

The observation of transgenic animals which contain the test vector and which increase stable circulatory human factor IX mRNA levels, as compared to the circulatory mRNA levels in transgenic animals which contain the control vector, demonstrates that the combination of AE5' and AE3' functions in increasing the activity of the exemplary CMV promoter.

#### **EXAMPLE 9**

### Liver-Specific Expression Of The Exmplary Human Factor IX Gene Under The Control Of The CMV Promoter

This Example investigates whether the presence of AE5' imparts liver specific activity to the CMV promoter, which otherwise drives gene expression in several tissues in addition to the liver.

The -802FIXm1 vector which contains AE5' and lacks AE3' (Figure 1) is used to construct a test vector in which the human factor IX promoter sequence of the -802FIXm1 vector is replaced with the CMV promoter sequence. This test vector is used in parallel experiments with the control vector of Example 8 in which expression of the hFIX gene is under the control of the CMV promoter in the absence of both AE5' and AE3'. Northern blot analysis is carried out as described *supra* (Example 3) in transgenic mice carrying the control vector or test vector. Animals expressing hFIX at high level are sacrificed at one month of age and total RNA is extracted from liver, lung, intestine, muscle, kidney, brain and heart and from untransfected HepG2 cells (negative control). The levels of hFIX mRNA in the different tissues are compared. It is expected that transgenic animals harboring the control vector will express hFIX mRNA in liver as well as in at least one other tissue. In contrast, the observation that transgenic animals which harbor the test vector express hFIX mRNA in the liver and not in other tissues indicates that AE5' confers liver-specific activity to the exemplary CMV promoter.

From the above, it is clear that the invention provides methods for age-related and liver-specific gene expression and models for age-related and liver-specific diseases.

25

30

5

10

15

20

#### **EXAMPLE 10**

# Construction Of A Series of Exemplary Human Protein C Minigene Expression Vectors

This Example and the following Examples 11-12 were carried out to further demonstrate the universality of the function of AE5' and AE3" sequences with respect to regulating expression of the exemplary human protein C as described in Example 7, *supra*. The human protein C genomic sequence has been previously reported (GenBank accession

number M11228; Figure 12B]. In particular, Figure 14 shows the nucleotide sequence (SEQ ID NO:85) of the 5'-end of the human protein C gene [Miao et al. (1996) J. Biol. Chem. 16:9587-9594]. In Figure 14, bases are numbered relative to the major transcription start site (+1) as previously described [Miao et al. (1996) J. Biol. Chem. 16:9587-9594]. Two minor start sites are marked with double asterisks. Exons are underlined. The translation start codon (ATG) is shown in boldface.

5

10

15

20

25

30

Using this sequence, eight protein C minigene expression vectors were prepared as shown in Figure 15. Fat lines at 5' and 3' ends represent flanking sequences. Introns are shown by blank rectangles. Exons are shown by shaded rectangles (corresponding to the 3' UTR part of the last exon). Arrows indicate transcriptional start sites. The vectors contained the human FIX AE5' sequence (SEQ ID NO:1; nt -802 through -771 of Figure 8; 32 bp in size) linked to Sph I linker sequences and/or the human FIX AE3' sequence (SEQ ID NO:93; nt 32,110 through nt 32,263 of Figure 8, 154-bp long) linked to Sse8387 I linker sequences. All the amplified sequences of the human protein C minigene vectors were verified by dideoxy sequencing.

The first human protein C (hPC) minigene vector (-1462hPCm1) was composed of the 5' flanking sequence of human protein C up to nt -1462, exon I sequence, the complete first intron at the natural site (1431 bp in length) and the contiguous following sequence containing exons 2-9 (derived from the hPC cDNA) and the 3' immediate flanking genomic sequence through nt 11,108 (325 bp in length). As shown in Figure 15, nucleotide 67 was the first base of intron I, nucleotide 1,497 was the first base of exon 2 (the first base of the Met codon is nucleotide 1,514), and nucleotide 10,488 was the last base of the stop codon.

To construct -1462hPCm1, a region spanning nt -1462 through 1,560 (3,022 bp in length) of the hPC gene was amplified by PCR (Expand High Fidelity PCR System, Boehringer Mannheim) using 5' and 3' primers containing Sph I linker and the unique internal Msc I site in the exon 2, respectively, and human genomic DNA as the template. The generated fragment, containing the 5' flanking sequence, exon I, intron I and a short 5' portion of exon 2 to the internal Msc I site, was then inserted into a hPC cDNA plasmid, PUC119-hPC, in between Sph I and Msc I sites by replacing its 5' portion of the hPC cDNA sequence. The 3' end of the resulting minigene, -1462hPCm', contained the entire 3' UTR, but only up to poly(A) attachment site (nt 10,783). Its 3' end region (the 3' sequence beyond the internal Sse8387 I site in the 3' UTR) was then freed by Sse8387 I/EcoR I

double digestion, and replaced with a Sse8387 I/EcoR I fragment (612 bp in the length, spanning nt 10,497 through 11,108 of hPC gene), which was generated by PCR. The hPC minigene, thus constructed was named -1462hPCm1, (approximately 4981 bp in length) and served as a parent construct for generating other hPC minigene constructs for generating other hPC minigene constructs as described below. This hPC minigene vector was used to construct the first transgenic mouse colony, which was used as a positive control for hPC expression.

5

10

15

20

25

30

The second human protein C minigene vector (-82hPCm1) was another control human hPC minigene. It was the same as -1462hPCm1, except that its promoter region was extended only up to nt -82 instead of to nt -1462, and thus contained the hPC upstream sequence (SEQ ID NO:86) from nt -82 to nt +1. Minigene -1462hPCm1 was subjected to Sph I digestion, followed by partial MscI digestion, releasing the 5' half region spanning nt -1462 in the 5' upstream through nt 1,547 at the internal Msc I site in exon 2. Due to another Msc I site in the first intron, partial digestion was needed to get the needed 5' end half fragment. This region was then replaced by smaller Sph I/Msc I fragments generated by PCR, spanning nt -82 in the 5' upstream through the internal Msc I site, thus generating -82hPCm1.

The third human protein C minigene vector (AE5'/-1462hPCm1) was the same as -1462hPCm1, except that it had an AE5' sequence (32 bp in length) inserted at the 5' end at an Sph I site.

The fourth human protein C minigene vector (-1462hPCm1/AE3'') was the same as -1462hPCm1, except it had an AE3'' sequence inserted at the Sse8387 I site within the 3' untranslated region (UTR) of the hPC minigene.

The fifth human protein C minigene vector (AE5'/-1462hPCm1/AE3'') was the same as -1462hPCm1, except it had both AE5' and AE3''. The AE5' fragment (32 bp in length, spanning nt -802 through -771 of hFIX gene) was amplified by PCR (Expand High Fidelity PCR System, Boehringer Mannheim) with human genomic DNA as template and PCR primers containing Sph I linker sequences. The AE3'' fragment spanning nt 32,110 through nt 32,263, and containing the potential stem-loop structure forming region (nt 32,142 through 32,243) was produced by PCR using primers containing the internal Sse8387 I site sequences. AE5' and AE3'' fragments with Sph I or Sse8387 I sticky ends, respectively, were then inserted into -1462hPCm1 at Sph I site at the 5' end and Sse8387 I site in the 3'

UTR, respectively, thus generating minigene AE5'/-1462hPCm1/AE3''. Therefore, AE5'/-1462hPCm1/AE3'' is the same as -1462hPCm1 except it has both AE5' and AE3' sequences of the hFIX gene at the 5' end Sph I site and Sse8387 I site in the 3' UTR, respectively.

The sixth and seventh human protein C minigene vectors (-849hPCm1 and -802hPCm1) were the same as -1462hPCm1, except that their 5' end sequences extended to nt -849 and -802, respectively, instead of to -1462. Construction of these minigenes was essentially the same as construction of the second human protein C minigene (-82hPCm1); minigene -1462hPCm1 was subjected to Sph I digestion, followed by partial MscI digestion, releasing the 5' half region spanning nt -1462 in the 5' upstream region through nt 1,547 at the internal Msc I site in exon 2. Due to another Msc I site in the first intron, partial digestion was needed to get the needed 5' end half fragment. This region was then replaced by smaller Sph I/Msc I fragments generated by PCR, spanning nt -849, or -802 in the 5' upstream through the internal Msc I site, thus generating -849hPCm1 and -802hPCm1, respectively.

The eighth human protein C minigene vector (AE5'/-82hPCm1) was generated by inserting the AE5' sequence with Sph I sticky ends into the second vector, -82hPCm1, at the 5' end Sph I site.

20 EXAMPLE 11

5

10

15

25

30

# Transient Expression of Exemplary Human Protein C Minigene Expression Vectors In Vitro In Human Hepatoma HepG2 Cell Line

Transient expression activities of the hPC minigenes of Example 10 were tested with HepG2 cells as described in Example 2, *supra*. Figure 16 shows the relative *in vitro* transient expression activities and standard deviations of the human protein C minigene expression constructs relative to the activity (approximately 70 ng/10<sup>6</sup> cells/48 hr, defined as 100% activity) of the control -1462hPCm1 construct.

The relative transient expression activities in Figure 16 show that the results obtained with AE3" when in tandem with the hPC gene are fully consistent with what the inventors previously observed with AE3" when in tandem with the hFIX gene (Figure 1); the presence of AE3" showed approximately a 30% suppression in transient expression in comparison to the minigenes which lacked AE3" (i.e., -1462hPCm1/AE3" compared with

-1462hPCm1, and AE5'/-1462hPCm1/AE3'' compared with AE5'/-1462hPCm1). The control constructs -1462hPCm1 and -82hPCm1 showed similar transient expression activities to each other.

As discussed above, these results were surprising because they were contrary to those previously reported by Miao et al. (1996), *supra*, when using a heterologous reporter gene, chloramphenical acetyltransferase (CAT), under the transcriptional control of varying lengths of the protein C 5'-end sequences.

5

10

15

20

25

30

#### **EXAMPLE 12**

### Generation And Analysis Of Transgenic Mice Harboring Human Protein C Minigene Expression Vectors

Transgenic animals were constructed using each of the eight expression plasmids described above in Example 10 according to standard methods [Hogan et al. (1994) in "Manipulating the Mouse Embryo, a Laboratory Manual" (Cold Spring Harbor Press, New York, 2nd Edition) as described in Example 3, *supra*. Fertilized eggs of mice were microinjected with the minigene transgene DNA and implanted into foster mother animals.

Circulatory hPC levels were monitored during longitudinal analyses of transgenic mice harboring the -1462hPCm1, -82hPCm1, and AE5'/-1462hPCm1/AE3'' minigene transgenes. At various ages, starting at one month of age, transgenic mice were individually subjected to blood sample collection via tail-tip snipping, and the obtained serum was routinely used to quantify hPC levels in the circulation using ELISA for each age point. The ELISA assay employed a mouse monoclonal anti-hPC antibody (Celsus Laboratories) as a first antibody and a rabbit polyclonal anti-hPC antibody (Celsus Laboratories) as a second antibody. Pooled human plasma (George King Bio-Medical) was used to prepare a hPC standard curve for each assay. The results are shown in Figure 17. The labeling in Figure 17 reflects the tag numbers of animals containing each minigene construct. Figure 17 shows representative animals with -1462hPCm1 (A), -82hPCm1 (B), and AE5'/-1462hPCm1/AE3'' (C) expression vectors.

Importantly, Figure 17 shows that age-regulation patterns were remarkably similar among all animals for each specific construct. In particular, the results show that transgenic animals containing the -1462hPC m1 construct contained age-stable levels of human protein C, *i.e.*, the animals expressed relatively constant levels of human protein C at different time

points during the life span of the transgenic animals (Figure 17A). In direct contrast, the presence of AE5' and AE3' sequences resulted in increased expression levels of human protein C over time (Figure 17C). These results confirm the universality of the function of AE5' and AE3' sequences in regulating expression of operably linked genes in an age-related manner.

5

10

15

The data also shows that, whereas transgenic animals containing the -1462hPC m1 construct exhibited relatively constant and relatively high levels (from about 100 to about 3000 ng/ml) of human protein C over time (Figure 17A), in dramatic contrast, transgenic animals containing the -82hPCm1 construct exhibited relatively low levels (from about 5 to about 40 ng/ml at 1 month of age) of human protein C which declined at a precipitous rate over time. Indeed, by the age of 5 months, human protein C levels were undetectable in all transgenic animals harboring the -82hPCm1 construct. These results demonstrate that the nucleotide sequence from nt -1462 to nt -83 of the human protein C gene directs age-stable expression as well as relatively higher levels of expression (as compared to the levels in the absence of the nucleotide sequence from nt -1462 to nt -83) of operably linked sequences of interest.

#### **CLAIMS**

1. A recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3.

### 2. A method, comprising:

5

10

15

- a) providing: i) a cell, ii) a nucleic acid sequence of interest, iii) a promoter sequence, and iv) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3;
- b) operably linking said nucleic acid sequence of interest, said promoter sequence, and said one or more age regulatory sequences to produce a transgene; and
- c) introducing said transgene into said cell to create a treated cell under conditions such that said nucleic acid sequence of interest is expressed in said treated cell.
- 3. A substantially purified nucleic acid sequence comprising at least a portion of SEQ ID NO:93.
- 20 4. The nucleic acid sequence of Claim 1, wherein said portion has age-related regulatory activity.
- 5. The nucleic acid sequence of Claim 1, wherein said portion is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO

WO 00/75279



NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144.

6. The nucleic acid sequence of Claim 3, wherein said portion is SEQ ID NO:91.

5

7. A recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof.

10

8. The expression vector of Claim 7, , wherein said nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase.

15

9. The expression vector of Claim 7, wherein said promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter.

20

25

30

10. The expression vector of Claim 7, , wherein said portion of SEQ ID NO:93 is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123; SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144.



- 11. The expression vector of Claim 10, , wherein said portion is SEQ ID NO:91.
- 12. The expression vector of Claim 7, further comprising in operable combination an age-related regulatory sequence selected from SEQ ID NO:1 and portions thereof.
  - 13. A host cell containing the recombinant expression vector of Claim 7.
- 14. The host cell of Claim 13, wherein said host cell is comprised in a tissue or organ in a living animal.
  - 15. The host cell of Claim 13, wherein said host cell is a gamete.
- 16. The host cell of Claim 13, wherein said host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.
  - 17. A method for expressing a nucleic acid sequence of interest, comprising:
  - a) providing:
    - i) a cell;
    - ii) a nucleic acid sequence of interest;
    - iii) a promoter sequence; and
- iv) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof;
- b) operably linking said nucleic acid sequence of interest, said promoter sequence, and said age-related regulatory sequence to produce a transgene; and
- c) introducing said transgene into said cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in said treated cell.
- 18. The method of Claim 17, wherein said treated cell is comprised in a tissue or organ in a living animal.

5

10

15

20

25

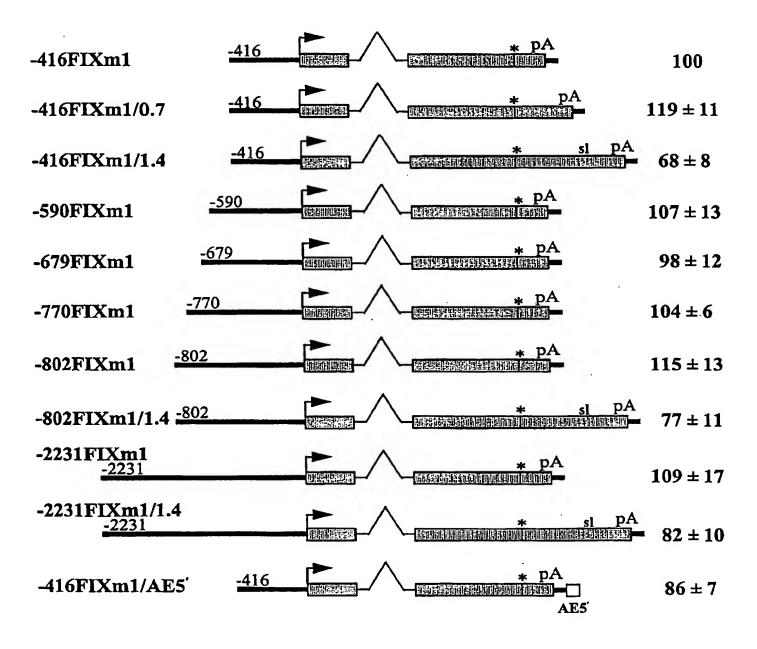
- 5
- 20. A method for expressing a nucleic acid sequence of interest, comprising:
- a) providing:
  - i) a cell;
  - ii) a nucleic acid sequence of interest;
  - iii) a promoter sequence; and

10

15

- iv) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, said nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90;
- b) operably linking said nucleic acid sequence of interest, said promoter sequence, and said nucleotide sequence to produce a transgene; and
- c) introducing said transgene into said cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in said treated cell.

Figure 1



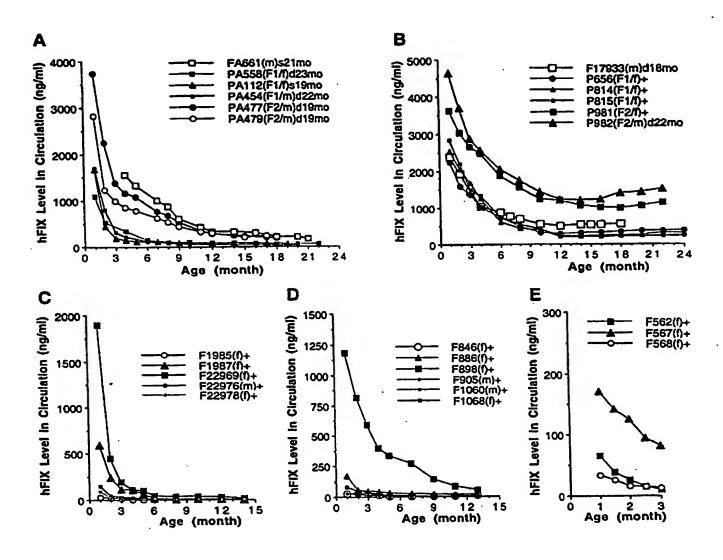
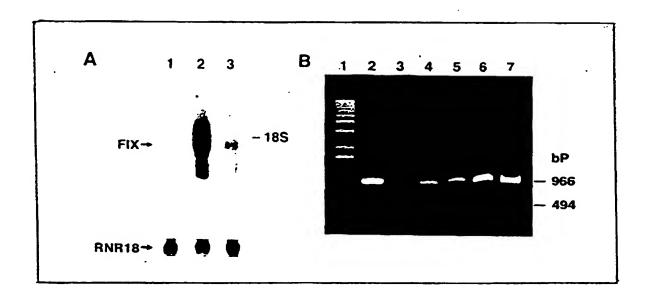


Figure 2

Figure 3



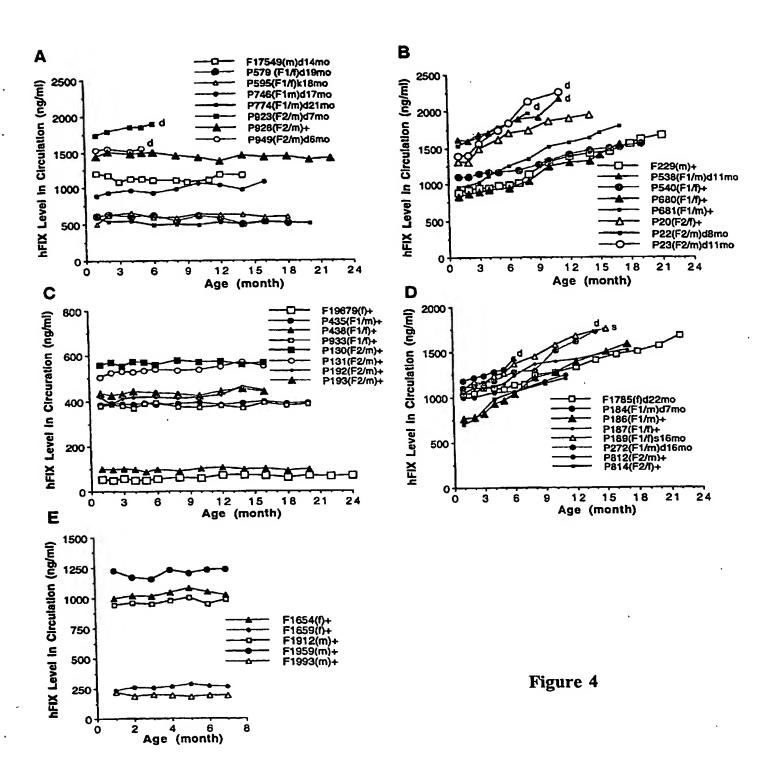


Figure 5

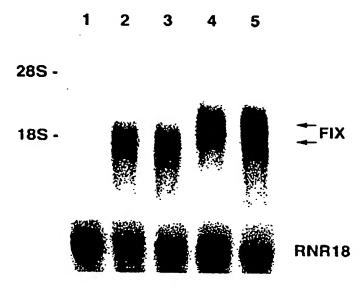


Figure 6

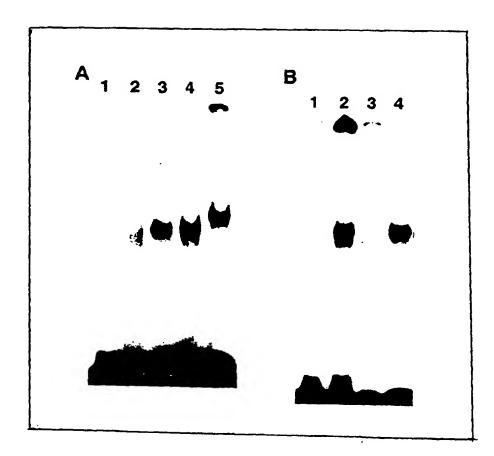
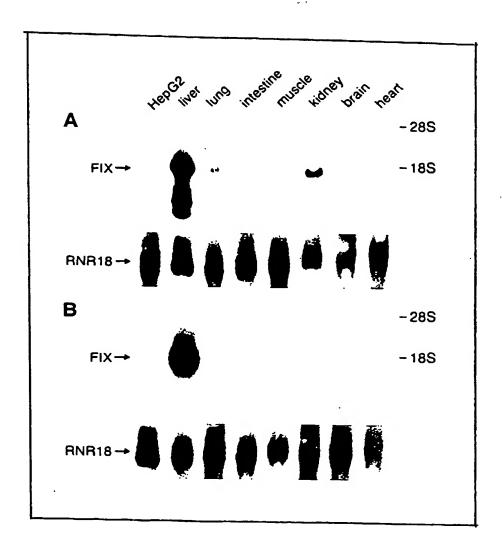


Figure 7



### Figure 8A

GTATATCTAG ACAGGCATTC						CICCIPICE	BC1.CC1.E1.1.C			
CELCETACEL .										-2866
										-2766
										-2666
										-2566 -2466
										-2366
										-2266
										-2166
										-2066
										-1966
AATTTTTGCA .	TATGALCAGA	CACTECTORGG	ALCALCACA	AGAATCTAAA	ATGAACTCAA	ACAAATTTAC	AGAMAAAAC	MACAACCCC	ATCAACAAGT	-1866
		CACTTCTCAA								-1766
										-1666
CAGCCATCCC	ATTATTGGGT	ATATACCAAA	GTATTATAAA	TCATGCTGCT	ATABACACAC	ATTECTOR	ACCOUNTED	CCTAGAAATA	CCATCTGACC	-1566
										-1466
GAGTTCATGT	CCTTTGTAGG	GCATGGATGA	AGCTAGAAAC	CATCATTCTC	AGCANACTAT	CCCAAGGACA	ALLACCALA	CACCOCCATA	AAAAGAAAT	-1366
										-1166 -1066
										-956
										-866
										-766
										-666
										-566
AGGCTGGAGA	CAATATATT	AACAAATAAA AATOCACGTA	CTGGACTACA	CLERCHE	AAAATTCTAC	TCTGTGAUAG	ACCTANTIAN	GAGGACAAAA	GACAAGCTAC	-466
ACTAACTGGA	CCACTCATAC	ATTGCTGATG	CALATOTAL	COCCOLCACA	TAGAATATAT	AAACAACCTT	ANGANTCIGA	CAGTAAAAAA	AAAAAATCAG	-366
CATGAACTGT	GCTGCCACAG	TAXATGTAGC	CACTATOCCO	PACAL CARC	TCALITICUTA	BCACCERCORC	CTCTCTGACA	AAGATACGGT	GGGTCCCACT	-266
AGCCCACGAA	ATCAGAGGTG	AAATTTAATA	ATGACCACTG	CCCATTCTCT	TOWNSHIELD	CAAGAGGCCA	TTCCABACIC	MATCHGCCA	CAGTGGCAGA	-166
					1000110100	CHICKOCCA		reconstance		-66
								•	-45 Met	
GATGGACATT	ATTTCCCAGA	AGTAAATACA	<b>GCTCAGCTTG</b>	TACTTTGGTA	CAACTAATCG	ACCTTACCAC	TTTCACAATC	TGCTAGCAAA	GCTT ATC	32
				-						32
Gln Arg Val	Asn Not II	e Net Ala (	Glu Ser Pro	Gly Leu Il	e The Ile C	ys Leu Leu (	ly Tyr Leu	Leu Ser Ala	Glu Cvs	
cue cae ata	AAC ATG AT	C ATG GCA	GAA TCA CCA	GGC CTC ATY	C ACC ATC TO	CC CTT TTA	GÁ TẤT CTA	CTC AGT GCT	CAN TOT	113
-10		•	•							
The										
ww	0111001111	TTAAAATACA	TIGACTATCC	TIGCCTTTIA	GATATAGAAA	TATCTGATGC	101C11C11C	ACTAAATTIT	CATTACATGA	211
TTTGACAGCA	ATATTCAACA	CTCTAACAGC	CACCACCCAC	-						
TAGATCGACA	ATGCTTATGA	TGCATTTATG	TTTARTARA	ACTOUTANGE	ACTOGITCTT	CTCLECTAG	GITTICITCI	TCITCATTIT	TAMAACTAAA	311
IAAAAAAA	AAAAAIIGGGA		ATAGCAGAAT	ATACTGARA	ABBATARCCE		CTTTCCACTT		110011100	411
CULANCANAA	GCACAAACAA	TGGCCTTATT	TACACAAAAA	GTCTGATTT	BAGATATATC	ACATTTCAAG	CTTTCACAAC	TATCTARGE	CC-CC-CC-CC-CC-CC-CC-CC-CC-CC-CC-CC-CC-	511 611
100111111	<b>WATTUTUTUT</b>		AAGTTTTAGT	TAXABACATAA	ACITTALCCT	TTCATTACCA		T18C1CC111		711
WILLY	AANICATIT	GICICIAIGI	CAAACATCTT	GGAGTTTGAT	TATTTCCCCA	BACACAATAC	TO A COTTO A COT	TCCCTACCCC	BCBBBBCCBB	811
CTINGONI.	TUNCATARAGE	AGTAGGAAGT	TAGCTAATGC	AACATATATC	ACTORCATOR	TTCACAACTA	CACTCACTOR		CACECCAACC	911
CATA CURRENT	MOVIMIL	CCATTTGGAL	AAACAGCATG	TTCTCACAGG	AAGCATTAT	CACACTTACT		CTACAATCAA		1011
TOACAGTACC	AGGATCAGGG	GTGCCAACCC	TAAGCACCCC	CAGAAAGCTG	ACTGGCCCCTG	<b>ずななすずなくといると</b>	TOCAGACATO	ATCTCACCOC	BC1110C1C1	1111
CTGAAATGCT	COLLOCOCA	TAAAAAAA	TACAGAGGTT	AAACTAGCAA	AGTGAGTAAA	GTCAAGGGAT	AAAGAAAATT	TCTTGGAAAA	CTCACAAAGC	1211
CCTTCTA & CA	COUNTY	TAGATATATC	TCATTAGTGT	.GACATCTGGG	AGGACAAAGC	ATCCANACCC	TTTCTTCTAT	ATAACTCCTC	1010010011	
ANT TOT UNION									WOULD FOUND	2311
CTAGCGATAA	ACCTGARGGG	TACTTALCTA	CITCAGATGC	TOCCOLLIC		ATALCCINCL	10000000	BCC1000000	~	. 1411
CINCONIAN	ACC TURNOUS	AAGITAAGIA	TACGATCCCC	TGGGGAAAGG	ATAGATAAGA	ATANGGATGA	ACCTGGCTTT	TGGAGCCTGG	GAAATAATGA	. 1411 1511
CCCTTGAGGA CGCTGAGTGA	AGGGGCCAGG TTTGCTGAGA	GGAATTTTTC TGTTTGCATT	TANGGATAGA	AGATAATACT CAGTATTAAT	ATAGATAAGA AAGGAGAAAG GCAGCACTCT	GCAATGTGAT CTTCTGCTAT	ACCTGGCTTT TCTGCAGCCA TAXACTCTCA	TGGAGCCTGG TTGTAGCCAG TTGGCTTCTA	GAAATAATGA AGATAATAAG AAAGGAGTTT	1411 1511 1611
CCCTTGAGGA CGGTGAGTGA CAGTTTAGTT	ACCTOANGE AGGGGCCAGG TTTGCTGAGA TTGTAAAGTG	GGAATTTTC TGTTTGCATT TATGCATCAA	TANGGATAGA TTANGGATAGA TTCATGCTGC AGATGTCCTT	AGATAATACT CAGTATTAAT TGCCTTTAGG	ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC	ATANGGATGA GCANTGTGAT CTTCTGCTAT ANCAGTTTGG ACACTAGGG	ACCTGGCTTT TCTGCAGCCA TANACTCTCA ANTITTGAAA PGCCAGCTAC	TIGAGECTICA TIGAGECTA TIGAGECTICA TITAAAACAGT	GAAATAATGA AGATAATAAG AAAGGAGTTT TCTGTAAAAC	1411 1511 1611 1711
CCCTTGAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA	AGGGGCCAGG TTTGCTGAGA TTGTAAAGTG ACCAGACTCC	GGAATTTTC TGTTTGCATT TATGCATCAA CTCTTTGATC	TANGGATCECC TANGGATAGA TTCATGCTGC AGATGTCCTT TANAGCAGCA	TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG	ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG	ACCTGGCTTT TCTGCAGCCA TAAACTCTCA AATTITGAAA TGGCAGGTAC	TGGAGCCTGG TTGTAGCCAG TTGGCTTCTA TTAAAACAGT TGTGTCAGGG	GAAATAATGA AGATAATAAG AAAGGAGTTT TCTGTAAAAC TACTAGGGGT	1411 1511 1611 1711 1811
COTTIGAGGA COTTGAGTGA COGTGAGTGA CAGTTTAGTT ATGGGGGATAA ATAGAGAAAG	AGGGGCCAGG TTTGCTGAGA TTGTAAAGTG ACCAGACTCC GAACACATTA	GGAATTTTC TGTTTGCATT TATGCATCAA CTCTTTGATC AATGGGGAAA	TANGGATUCCC TANGGATAGA TTCATGCTGC AGATGTCCTT TANAGCAGCA CANTTGATAG	TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG	ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT	ACCEGNATE ACCEGNATION AND ACCEGNATION AND ACCEGNATION AND ACCEGNATION AND ACCEPTANCE AND ACCEPTA	TIGAGECTOG TTGTAGECAG TTGGETTCTA TTAAACAGT TGTGTCAGGG GCACTAGGTA	GAAATAATGA AGATAATAAG AAAGGAGTTT TCTGTAAAAC TACTAGGGGT CTAAGGGATC	1411 1511 1611 1711 1811 1911
CINCENTAN CCCTTGAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAG CTGATATAAG	ACCIONAGE AGGGGCCAGA TTTGCTAAAGTG ACCAGACTCC GAACACATTA GGCATTTTAT	GGAATTTTC TGTTTGCATT TATGCATCAA CTCTTTGATC AATGGGGAAA GCAAAGAAGA	TACGATECEC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCATTCGTG	TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAAATA CAAAGACTC	ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT TTTTCCTAG	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA	ACCTGGCTTT TCTGCAGCCA TANACTCTCA AATTITGAAA TGGCAGGTAC GATAAAATGT GATGAGTAGC	TEGASCETEG TTGTAGECAG TTGGETTETA TTAANACAGT TGTGTCAGGG GCACTAGGTA CGTTCTCTCT	GANTANTON AGATANTANG ALAGGAGTT TCTGTANAAC TACINGGGGT CTANGGGATC CTITANAIGT	1411 1511 1611 1711 1811 1911 2011
CCTTGAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAG CTGATATAAG ACTATGACAA	ACCIONAGE TITGCIGAGA TITGCIGAGA TITGIAAAGIG ACCAGAATIA GAACACAITA GGCATITIAT GTGAGACAGG	AGGITAGTA GGAATTTTC TGTTTGCAT TATGCATCAA CTCTTTGATC AATGGGGAAA GCAAAGAAG TAAACTAGGC	TACGATCECC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCACTCGTG AGAGCTGGTC	TGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGGAATA CAAAGACTCA ATCAGATAAT	ATAGATAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT TTTTCATCTG GCTTTGCAAG GAGTCATTA	ATMAGGATGA GCAATGTCAT CITCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTGAGGT	ACCIGATITI TOTOCAGCCA TANACTOTICA ANTITIGAA TOGCAGGIAC GATANAATGT GATGAGTAGGA ATTICAGGAG	TEGASCETEG TIGTAGECAG TIGGETTETA TIANACAGT TETGTEAGGG GEACTAGGTA CGITCTETET TITTGTATEG	GANTANTON AGATANTANG ALAGGAGTTT TCTGTANAAC TACINGGGGT CTANGGGATC CTITANATGT TTCCATATGG	1411 1511 1611 1711 1811 1911 2011 2111
CCCTTCAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAG CTGATATAAG ACTATCACAA CGAAGTATAA	ACCIOANGIGA AGGGGCCAGG TTTGCTGAGA TTGTAAAGTG ACCAGACTCC GAACACATTA GGCATTTTAT GGCATTTTAT GTGAGACAGG ACATAAAGGAG	ACTITAGES GGAATITTC TGTTTGCATT TATGCATCAA CTCTTTGATC AATGGGGAAA GCAAAGAAGG TAAACTAGGC TACCACTGAT	TACGATECEC TANGGATAGA TICATGCTGC AGAIGTCCTT TANAGCAGCA CANTTGATAG ATCACTCGTG AGAGCTGGTC AGAGCTGGTC	TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGAATA CAAAGACTCA ATCAGATAAT	ATMATAMA AAGGAAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT TTACTGAG GCTTTGCAAG GCAGTCATTA	ATMAGGATGA GCAATGTGAT CITCTGCTAT ACAGTTTGG ACACTACGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTAAGGAG	ACCIGGETTT TCTGCAGCCA TAMACTICTCA AATTITGAAA TGGCAGGTAC GATAAAATGT GATAGAGTAGG ATTTCAGGAG ATTTGACAAT	TEGAGCCTGG TTGTAGCCAG TTGTAGCCAG TTGAAAACAGT TTTAAAACAGT TGTGTCAGGG GCACTAGGTA CGTTCTCTT TTTTGTATGG AMATGCAAT	GANTANTGA AGATANTAAG AAAGGAGTTT TCTGTAAAAC TACTAGGGGT CTTAAAGGT TTCCATATGG ATGGAGGTAT	1411 1511 1611 1711 1811 1911 2011 2111
CENTRAGA CEGTGAGTGA CAGTTTAGTT ATGGGGATAA ATACAGAAAG CTGATATAAG ACTATGACAA CGAAGTATAA CFAGGAGAGA	ACCIONAGE TITECTEAGA TITECTEAGA TICTAAAGTG ACCAGACTCC GAACACATTA GGCATTTTAT GTGAGACAGG ACATAAGGAG GCATTCCTGA GCATCCTGA	AGGITAGTA GGAATITITC TGTTTGCATT TATGCATCAA CTCTTTGATC AATGGGGAAA GCAAAGAAGG TAAACTAGGC TACCACTGAT GACTATTGCA	TACGATECEC TANGGATAGA TICATGCTGC AGATGTCCTT TANAGCAGCA CANTTGATAG ATCACTCGTG AGAGCTGGTC GGCTGATTTA ATTATCTGCG	TGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGAATA CAAAGACTCA ATCAGATAAT GGATGCCCAG	ATMGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT TTTCATCTG GCTTTGCAAG GAAGTCATTA TCTGGCAACA	ATMAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTAAGGAG CGGTAATGAA CTGAACGGTA	ACCTGACTT TCTGCAGCA TAAATTTGAAA TGCCAGGTAC GATAMATGT GATGAGTAGG ATTTGAGGAG ATTTGAGAAA AFGATAGTGG CCAGTACAAA	TEGAGCCTGG TTGTAGCCAG TTGTAGCCAG TTAAAACAGT TGTGTCAGGG GCACTAGGTA CGTTCTCTT TTTTGTATGG AAAATGCAAT GGGAGGGGGCCC	GANTANTGA AGATANTAAG AAAGGAGTTT TCTGTANAAC TACTAGGGGT CTTANATGT TTCCATATGG ATGGAGGTAT CGTACCAAGA	1411 1511 1611 1711 1811 1911 2011 2211 2211 2311 2411
CCTTCAGGA CGGTCAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAG CTCATATAAG ACTATCACAA CGAAGTATAA CFAGCAGAGA ATATTAAGGA	ACCIOANCE TITGCIGAGA TITGCIGAGA TITGCIGAGACTCC GAACACATTA GGCATTITAT GTGAGACAGG ACATAAGGAG GCATCCIGA AGTAGAAGTG	AGGITAGTA GGAATTTTC TGTTTGCAT TATGCATCAA CTCTTTGATC AATGGGAAA GCAAAGAAGG TAAACTAGGC TACCACTGAT GACTATTGCA GTATGACTTA	TACGATECEC TANGGATAGA TTCATGCTGC AGATGTCCTT TANAGCAGCA CAATTGATAG ATCACTCGTG AGAGCTGGTC GGCTGATTTA ATTATCTGCG ACCACTCGCC	TEGGRANAGE AGATAATAAT CAGTATTAAT TEGCTTTAGE CATTCAGACA TEAGGCCAGG AGAGAGATAA CANAGACTCA ATCAGATAAT GGATGCCCAG GGAGACATAA	ATAGATAGA AAGGAGAAAG GCAGCACTCT TTATTATTACCC TTACTGAGGTTT TTATCACTG GCTTTGCAAG GAAGTCATTA TCTGGCAACA AGGCTAGAAC	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTAAGGAG CCCTAATGAA CTGAACAGTA	ACCTGGCTT TCTGCAGCCA TAMÀTCTCA AATTTGAAA TGCCAGGTAC GATAMATGT GATGAGTAGG ATTTCAGGAG ATTTCAGCAAT ATGATAGTGG CAGTACAAA	TEGAGECTEG TTGTAGECAG TTGGETTETA TTAAACAGT TETGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAAT GGGAGGGGCC AAAAGAGGGGC	GAMTANTGA AGATANTANA AAAGGAGTTT TCTGTANANC TACINGGGGTT CTTANANGT TTCCATATGG TTCCATATGG TTCCATATGA CGTACCANGA AGTTCANANG	1411 1511 1611 1711 1811 1911 2011 2111 2211 2311 2411 2511
CETTCAGGA CGGTGAGTGA CGGTGAGTGA ATGGGGATAA ATAGAGAAAAG CTGATATAAAG ACTATCACAA CGAAGTATAA CTAGCACAGA AIATTAAGGA AGAATTAAAGGA AGAATTAAAGGA AGAATTAAAGGA AGAATTAAAGGA AGAATTAAAGGA	ACCIONAGE TITGCIGAGA TITGCIGAGA TITGTAAAGTG ACCAGACTIC GAACACATTA GGCATTITAT GTGAGACAGG ACATAAGGAG GCAGTCCIGA AGTAGAGAAAC AGTAGAGAAAC AGTAGAGAAAC	AGUITAGIA GGAATITITC TGTTGCATT TATGCATCAA CTCTTTGATC AATGGGGAAA GCAAAGAAGG TAAACTAGGC TACCACTGAT GACTATTGCA GTATGACTTA TGCCTCTTTA	TACGATECCE TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCACTCGTC GGCTGATTTA ATTATCTGCG ACCACTTGCTG ACCACTTGCTG	TGGGGNAAGG AGATAATACT CAGTATTAAT TCCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGGATAAT ACAAGACTCA ATCAGATAAT GGATGCCCAG GGAGACATAA TATGGAAGG	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTAGC TTACTAGGTT TGAGAGGTTT TTTTCATCTI GCATTGCAAG GAAGTCATTA TCTGCCAACA AGGCTAGAAC GAATAGGCTA	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAMAA AACGTGAGGT ACCTAAGGAG CCCTAATGAA CTGAACAGTA GAGTCTTGGG GATCTTTGGG FAGTCTTGGG FAGTCTTGGG	ACCIGACIA TANACTOTA ANTITIGANA TOCCAGGIAC GATANANTOT GATGAGIAGA ATTOGACAA ATTOGACAA ACCITAGIAGA GAGIACAAA GACTITGIAGA	TEGRECTEG TTGGECTETA TTGGECTETA TTANACAGT TEGTECAGG GCACTAGGTA CGTTCTCTCT TTTGTTATGG AAAATGCAAT CGGAGGGGGG TGATGTGAT	GAMTANTGA AGATANTANG AMAGGAGTTT TCTGTANAGG TACHAGGGATC CTHANAGGT TTCCATATAGG ATGGAGGTAT CCTACCANGA AGTICANATG TATGGACCAC	1411 1511 1611 1711 1911 2011 2111 2211 2311 2311 2411 2511
CCOTTGAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAG CTGATATAAG ACTATGAGAA CGAAGTATAA GAAGTATAAA ATATGAGAA ATATGAGAA ATATGAGAA ATATGAGAA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTATGTTTCC	ACCIONADE TITGCTGAGA TITGCTGAGAC ACCAGACTCC GAACACATTA GGCATTTAT GTCAGACAGG GCAGTCCTGA ACGIAGAGAG GCAGTCCTGA AGTAGAGAC TCATATATAA	AGGITMGTA GGAATITTC TGTTGCATT TATGCATCA CTCTTTGATC AATGGGGAAA GCAAAGAAGG TAACCACTGAT TAACCACTGAT GACTATTGCA GTATGACTTA TGCGCTCTTA AAATAAGATG	TACGATUCCE TAAGGATAGA TTCATGCTGC AGATGCCTT TAAAGCAGCA ATCACTCGTG AGAGCTGGTC GGCTGATTTA ATTATCTGCG ACCATCTGGG GTCTGACTGG	TGGGMAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGAATAA ATCAGATAAT GGATGCCCAG GGAGACATAA TATGGAAGGG CAGAGTCCGA TATGGAAGGG CAGAGTCCGA TATGGAAGGG	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTAGC TTACTGAGTT TTAGAGGTTT TTTTCATCTG GCTTTGCAAG GAAGTCATTA TCTGGCAACA AGGCTAGAAC GAATTGCTA ATCCTGAATG	ATANGGATGA GCAATGATGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTCAGGT ACCTAAGGAG CCGTAATGAA CTGAACAGTA GAGTCTTGGG TTTTAGTATGG	ACCIGACITY TCTGCAGCCA TCAACTCTCA AATHTGAAA TGCCAGGTAC GATAMATGT GATGAGTAGG ATTCAGGAG ATTCAGGAG ATTGGACAAT AFGATAGTGG CCAGTACAAA CACTITGTGT TTACCTTGCA	TEGAGECTEG TTGTAGECAG TTGGETTETA TTAAAACAGT TGTGTCAGGG GCACTAGETA CGTTCTCTT TTTTGTATCG AAAATGCAAT GGGAGGGGG TTGATGTGAT AAGCCCTTAG	GAMTANTGA AGATANTAG AMAGGAGTTT TCTGTANAAC TACTAGGGGT CTTAANTGT TTCCATATGG ATGGAGGTAT CGTACCAAGA AGTTCAANTGA TATGGACCAC CCTGTATGAA	1411 1511 1611 1811 1911 2011 2211 2211 2311 2411 2511 2611 2711
CCOTTGAGGA CGGTGAGTGA CAGTTTAGGTTA ATGGGGATAA ATAGAGAAAC CTGATATAAG ACTATGACAA CTAGGGAGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTATGTTTCC GTCCAACACA TATGTGACTA	ACCIONAGE TITGCIGAGA TITGCIGAGA ACCAGACTCC GAACACATTA GGGATITITAT GTGAGACAGG ACATAAAGGAG GCAGTACCIGA AGTAGAAGT TCATITATA GCAGAACAC TCATITATAA GCAGAACAC TAGGAGGAT TAGGTGGAGT	AGGANTITTC GGTATGCATT PATGCATCAA CTCTITGATC AATGGGGAAA GCAAGGAGG TAAACTAGGC TACCACTGAT GACTATGCA GATATGCATTA TGCCTCTTTA AAATAAGATG TACATATACA TACATATACA TACATATACA TACATATACA TACATATACA TTTGGAACTC	TACGATUCCE TAAGGATAGA TTCATGCTGC AGATGTCTT TAAAGCAGCA CAATTGATAG AGAGCTGGTC AGAGCTGGTC GGCTGATTTA ATTATTGCG ACCATCTGGG GTCTGACTGC ACAGTGCCTA TTAACGTATTA	TGGGANAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAGGACATAA ATCAGATAAT GGATGACAA TATGGAAGG CAGGATCATA TATGGAGGC CAGGTTCGA TCTGGTGGGA CTGGCTACAT AAACAGTAGT	ATAGATAGA AAGGGAGAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT TTTTCATCTB GCATTGCAAG GAGTCATTA TCTGGCAAC GAATGGCTA ATCCTGAAG CTTTTGTGAG TATGATATG TATGATATA	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAMAA AACGTGAGGT ACCTAAGGAG CCGTAATGAA CTGAACGTA GAGTCTTAGGAG TTTTAGTATG AGTTAGGAG AGTTAGGAG AGTTAGGAG AGTTAGGAG AGTTAGGAG AGTTAGGAG AGTTAGGAG AGTTAGGAG	ACCTGGCTA TCTGCAGCA TAACTCTCA AATTTGAAA TGGCAGGTAC GATAAATGT GATGAGTAGG ATTGGGAAA ATTGGACAAT AFGATAGTGG CAGTACAAA GACTTGGTT TTACCTTGGT TTACCTTGGT TTGGAAAGAT TTGGAAT TTGGAAT TTGGAAT TTGGAAT TTGGAAAGAT TTGGAAT TTGGAAT TTGGAAAGAT TTGGAAT TTGG	TEGAGECTEG TTGTAGECAG TTGGTTCTA TTANACAGT TTGTGTCAGG GCACTAGGTA CGTTCTCTT TTTTGTATGG AAAATGCAAT CGGAGGGGGC TAAAGGGG TTGATGTATGATGATGATGATGATGATGATGATGATGATG	GAMTANTGA AGATANTAG ALAGGAGTTT TCTGTANAG TACTAGGGGTT CTTAGAGGTT TTCCATATGG ATGGAGGTAT CCTACCAAGA AGTTCAAATG TATGGACCAC CCTCTATGAA TCAGGACCAC TCAGGAGAGT TCAGGAGAGT	1411 1511 1712 1811 1911 2011 2211 2211 2411 2511 2611 2711
CCOTTGAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAG CTGATATAAA ACTATGACAA CGAAGTATAAA CTAGGAGAGA AIATTAAGGA TGTATCTTOC GTCCAACACA TATTTGAGTA TAGGAGTA	ACCIONADA TITGTIGAGA TITGTIANAGTG ACCAGACTCC GANCACATTA GGCATTITAT GGCAGACAGG ACATANGGAG GCAGTCCTGA AGIAGAGAC TINTATANAGGAGAC TANTATATANAGGAGAC TAGGGGAGCAT ACGTGGCAGTA	AGGITATITA GGAATITTA TATGCATCA CTCTTTGATC AATGGGGAAA GCAAAGAAGG TAACCACTGAT GACTATTGCA GTATGACTTA TGCGCTCTTTA AAATAAGATG TACATTACA TTTTTTTTTT	TAGGATUCCE TAAGGATAGA TTCATGCTGC AGATGATAG ATCATCTGTG AGAGCTGGTG AGAGCTGGTG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACATGCCTA TTAGCTATTA ATTATCTTCGT ACCATGCCTA ATTATCTTCGG ACCATGCCTA ATTATCTTCGTA ATTCTTCTGTA	TEGGOMANGE AGATANTACT CAGTATTAAT TECCTTTAGG CAITCAGACA TEAGGCCAGG ACAGAGATAA CANAGACTCA ATCAGATAAT GGATGCCCAG GGAGCATAAA TATGGAAGGG CAGAGTCTGA TCTGGTGGGA ANACAGTAGT ANACAGTAGT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT TTATCATCTG GCTTTGCAAG GAAGTCATTA AGGCTAGAAC AAGGCTAGAAC CAAATAGGCTA ATCCTGAATG CTTTTGTGAG TATCATATATAC TACAGCAGGT	ATANGGATGA GCANTGTGAT CTTCTGCTAT ACACTATCG ACACTACGG CAATATATGT GGTCTTAAAA AACGTCAGGT ACCTTAGGAG CCGTAATGAA CTGAACAGTA GAGTCTTGG GATCAGTAGA GTTTAGTATG GATCAGGAG GTTTGGGGG GTTTGGGGG GTTTGGGGG	ACCTGGCTT TCTGCAGGCA TAAACTCTCA TATTTGAAA TGGCAGGTAC GATAAAATGT GATCAGGTAG ATTTGACAAA TGATCAGGAC ATTTGACAAA TGATCAGGAC ACCTTGTGT TTACCTTGCA GATAAAGAT CTGGAGAGTA CTGGAGAGTA CTGGAGAGTA CTGGAGAGTA CTGGAGAGTA CTGGAGAGTA CTGGAGAGTA	TEGAGCCTEG TTGTGTCTA TTTAMACAGT TTGTGTCAGG GCACTAGGTA CGTTCTCT TTTTGTATGG AMATGCAAT GEGAGGGGG TTGATGTATA AGCCCTTAG AMACTGCA AMACTGCA AMACTGCA AMACTGCA AMACTGCA AMACTGCA AMACTGCA AMACTGCA CCAAAGTTAC CCAAAGTTAC CCAAAGTTAC AMACTGCATCA AM	GAMTANTGA AGATAMTANG ANAGGAGTTT TCTGTAMAGG TACTAGGGATC CTTTAMATGT TTCCATATGG ATTGAGGTAT TCTACCAGA AGTTCAMATG TATGGACCAC TCTGTATGAA TCAGCACAGT TCAGGAGAGGT TCAGGAGAGGT TCAGGAGAGGT	1411 1511 1711 1811 1911 2011 2211 2311 2311 2511 2611 2711 2811
CCOTTGAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAG CTGATATAAG ACTATGACAA ACTATGAGAGA ACTATGAGAGA ACTATTAAGG ATATTAAGG ATATTAAGGA ATATTAAGGA TGGATATGT TGGATCAGC TGTACTTCT TGTACTTGT TAGTACAGTG TATACTTGTA	ACCIONALA TITGCTEGAGA ACCAGACTCC GAACACATTA GGCATITTAT GTGGACACAG GCAGTCCTGA AGTAGAGGAG CTAAGAGAAC TCATTIATAA GCAGAGCAC TCATTIATAA GCAGAGCAG TAGGTGGGAG CTGGCAGTAGAGAC TAGGTGGAGAC	AGGITATITA GGAATITITA TATGCATCA TATGCATCA AATGGGAAA GCAAGGAGG TAAACTAGGC TACACTGAT GACTITA AATAAGATG TAGATTATACA TAGATATACA TAGATAGACT TAGATAGA	TACGATICECE TAAGGATAGA TTCATGCTGC AGATGTCTT TAAAGCAGCA ACACTCGTG AGAGCTGGTC AGAGCTGGTC ACCATCTGGG	TGGGANAGG AGATAATACT CAGTATTAAT TCCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGATAA ATCAGATAAT GGATGCCCAG GGAGACATAA TATGGAAGGC TCTCGTGGGA TCTCGTGGGA TTTTGTTGAC TTTTGTTGAC TTTTGTTGAC TTTTGTTGAC TTTTGTTGAC TTTTGTTGAC TTTTGTTGAC TTTTTTTTGAC TTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTAGCT TTACTAGGTT TTATTCATCTG GCATTCCAAG GCATTCCAAG GAAGTCATTA TCTGCCAACA AGGCTAGAAC GAAATGGCTA ATCCTGAATG CTTTTGTGAG TATGATATAC TACAGCAGT AGGCTAGAGG	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTAAMA AACGTCAAGGA ACCTTAAGGAG CGCTAATGAA CGTAACAGTA GAGTCTTGGG FATTTTGGG GATTAGGAG GATTAGGAG GATTAGGAG GATTAGGGAG GTTTAGGGAG GTTTAGGGAG GTTTAGGGAG CGCGACTAGA	ACCIGACIA TANACTOTA ANTITIGANA TOCCAGGIAC GATANATOT GATGAGTAG ATTOGACAT ATTOGACAT ATGATAGTAG CAGTIACAN GACTITIGT TIACCITIGA GATATGGAT TIACCITIGA GATATGGATG CTGGANAGAT CTGGAGAGTG CTGTGGANG	TEGROCTEGE TITGAGECAG TITGATTETA TITAMACAGT TOTGTCAGGG GCACTAGGTA CGTTCTCTT TITTGTATGG AAAATGCAAT CGGAGGGGG TITGATGTGATGGGGG TITGATGTGATGTG	GAMAMATGA AGAGAGATTT TCTGTAMAG TACTAGGGGT CTTAMATGT TTCCATATAG ATOGAGGTAT CGTACAAGG ATTGACAAGG TATGGACCAC CCTTATAAATGT TATGGACCAC TATGGACCAC TATGGACCAC TATGGACCAC TATGGACACAC TCAGGAGAGAG TCAGGAGACAC TCAGGAGACAC AGGGCATAAA	1411 1511 1711 1811 2011 2213 2211 2311 2411 2511 2611 2711 2812 2911
CINTEGRIGA COSTTERAGGA CAGTTERGTA ATREGGANAA CTATTACAG ACTATGACAA CHAGTATAAA CHAGTATAAA CHAGTATAAA ATATTAAGGA ATATTAAGGA ATATTATTCC CTCLAACACA TATTEGACTA TACTTGACTA TEALACTGTT TALACTTGTT	ACCIONADA TITGTIGAGA ACCAGACTICA GANCACATTA GGCATITTAT GTENGACAGG ACATANGGAG GCAGTCCTGA AGTAGAAGAG CTANGAGAG CTANGAGAG CTANGAGAG CTANGAGAG CTANGAGAG CTANGAGAG CTANGAGAG CTANGAGAG TANGTIGAGTA CCTGGCAGTA AGAGAGGATA	AGGITMETT GENTTECATT PATECATCA CTCTTTGATC ANTEGGRANA GCANAGNAGG TANCTAGGC TACCACTGAT GACTATGCA GTATGCATTA TGGCTCTTTA TGGCTCTTTA TACATATACA TACATATACA TACATATACA TACATATACA TACATATACA GTGGTGTTAT AGTGAGGACT GTGGTGGAGGCC GTGGAGGCCT	TACGATUCCE TAAGGATAGA TTCATGCTGC AGATGTCTT TAAAGCAGCA CAATTGATAG ACACTCGTG ACACTCGTG ACTCGTG ACTCGTGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGT TAACCTATAA ATTGTTCTGT GCTCAGTAAA TCTTAAATAT TGGGGAAAACCC	TGGGGNAAGG AGATAATACT CAGTATTAAT TCCCTTTAGG CATTCAGACA TGAGGCCAGG ACAGGACATAA ATCAGATAAT ATCAGATAAT TATGGAAGG CAGAGTCTGA TATGGAGG CAGAGTCTGA TATGGAGGG AAACAGTACA AAACAGTACA TATGTGAGGA TTTTGTTGAC TTTTCCCTAC	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTAGC TTACTAGGTT TGAGAGGTTT TTTTCATCTE GCATTGCAAG GAAGTCATTA TCTGGCAACA AAGGCTAGAAC GAAATGGCTA ATCCTGAATG CTTTTGTGAG TATGATATAC TACAGCACGT AGGCTCAGGG ATGCCTTAGT	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTAMAA AACGTGAGGT ACCTAAGGAG CCGTAATGAA CTGAACGTA GATTTAGGTA GATTTAGGTAG GATTAGGAG GTTTGGGCAT CCGGACTAGA TGGTCACACA	ACCTGGCTAT TANCTCTCA AATTTGAAA TOGCAGGTAC GATAAAATGT GATGGGTAGG ATTGGGCAA ATTGGACAAT AFGATAGTGG GCAGTACAAA GACTTGGTT TTACCTTGGT TTACCTTGGT TTGGAAAGAT TTGGAAAGAT CTGGGAAAGAT CTGGGAAAGT TTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGTGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGGAAAGTT CTGCAAAGTT CTGCAAAAT CTGCAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAT CTGCAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAAT CTGCAAAT CTGCAAAAT C	TEGAGECTEG TTGTAGECAG TTGGETTETA TTANANCAGT TEGTETCAGG GCACTAGETA CGITCTCTT TTTTGTATGG AAAATGCAAT CGGAGGGGG AAAAGAGGG TTGATGTGA AAGCCCTTAG AAGCCCTTAG AAGCAGGAAT ACATGGAAT ACATGGAAT ACATGGAAT ACATGGAAT ACATGGAAT ACATGGAAT	GAMTANTGA AGATANTAG AGAGGATTT TCTGTANAGC TACTAGGGGT CTTAGAGGT TTCCATATGG ATGGAGGTAT CCTACCAGG AGTTCAAATG TATGGACCAC CCTCTATGAA TCAGGACAGT TCAGGAGAGC TTCTGATAAC AGGGCATAAA ATATATAGGAGGG	1411 1511 1711 1811 1911 2011 2211 2311 2311 2511 2611 2711 2811
CCOTTGAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAG CTGATATAAG ACTATGACAA CGAAGTATAA ATAGAGAGA ATATGACAA CGAAGTATAAG ATATTGACAA ATATTGACAA ATATTTGACA ATATTTTCC GTCCAACACA TBATCTGACTA TAGTACATTGT TAGATAGTA TAGAAAAAGGA ACAGGAGTTA	ACCIONADA TITGTIGAGA ACCAGACTOC GAACACATTTA GCAGACTAGA GCAGTITTATA GCAGACAGA GCAGTOCTGA AGTAGAAGTG CTAAGAGAAC TCATITIATAA GCAGAAGCAT TAGGTGAGTA CCTGGCAGTA TAGGTGAGTA AAAAGCTGAG GAGAAAGCG	AGGINATITA GGANTITTA TOTTGATT PATGCATCA CTCTITGATC AATGGGGAAA GCAAAGAAGG TAACCACTGAT GACTATTGCA CTATGACTTA TGGACTTA TAACAATGAT TTGGAACTA TTTGGAACTA TTTGGAACTA AGTGAGGACC CCTGGAGGCC GGGAACCAGA	TACGATUCCE TAAGGATAGA TTCATGCTGC AGATGCCTT TAAAGCAGCA ATCATCGTG AGACTGGTG AGACTGGTG ACCATCTGGG GCTGATTTA ATTATCTGGG GTCGATCTGGG TTCGACTGG TTAGCTATTA ATTATCTTCGT TTAGCTATTA ATTATCTTCGT GCTCAGTGATAAATAT GCGTGAGAACG GTGAGAACGG	TEGGONANGG AGATANATACT CAGTATTAAT TGCCTTTAGG CAITCAGACA TGAGGCCAGG ACAAGACTCA ATCAGATAAT GCATGCCCAG GGAGACATAA TATTGAGAGG CAGAGTCTGA TCTTGTGGGA CAGAGTAGT TTTTGTTGAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TTACTGAGTT TTACTACTG GCTTGCAAG GCAGCACA AGGCTAGACA AGGCTAGACA ATCCTGAATG CTTTTTGTGAG TATCATATAC TACAGCACGT TACAGCACGT AGGGCTAGAG ATGCCTTGTT AGGGCCAGGG ATGCCTTGTT AGGGCCATGG	ATANGGATGA GCANTGTGAT ACACTATCG ACACTACGG CAATATATGT ACGTCAGGT ACCTAAGGAG ACCTAAGGAG CCGTAATGAA CTGAACAGTA GAGTCTGGG TTTTAGTATG ATTAGGAG GTGTGGGGAT ACTTAGGAG TTTTAGTATG ACTTAGGAG TTTTAGTATG ACTTAGGAG TGGTGGGAT ACGTAGA AGTTAGGAG AGTAGGAAA AGGAGAACA AAAAGGGAG AAAAAGGGAG	ACCTGGCTAT TCTGCAGGCA TAAACTCTCA AATTTTGAA TGGCAGGTAC GATAAATGT GATGAGTAGG ATTCGAGAA ATGGACAAT TGACATAGAGA ATGGACAAT TAACTTGAC AATAATGGTG CAGTACAAA GACTTGGACAGT CTGGAGAGTG CTGGAGAGTG CTGGGAAGAT TTGCCCAGC GAGTTAATCA AGATCACAA AGATCACAA	TEGAGCETEG TTGTGTCAGG TTGGTTCTA TTAMACAGT TTGTTCAGGG CCACTAGGTA COTTCTCT TTTTGTATGG AMATGCAAT GCGAGGGGC AMAGAGGGG TTGATGTATA AACCCCTTAG CCAAAGTTAC CAAAGGTTAC CAAGGCTGT CCTAGGCAGT ATTCAAAGTG	GAMTANTGA AGAGMATTAG AAAGGAGTTT TCTGTAMAGT TACTAGGGGT CTATAGGGGTTT TCCATCATAGGGTAT CCTACCAGA AGTTCAMATG TATGGACGATA TCAGCACCAC AGTCAMATG TCAGCACCAC TCAGCACCAC TCAGCACCAC TCAGGACACG TCAGGACAC TCAGGACACG TCAGGACAC TCAGGACACG TTAGGACAC TGAGATGGATAC AGGATGGATA TTTAGACCAC	1411 1511 1611 1711 1811 2011 2111 2211 2311 2411 2513 2611 2711 2811 2911 3011
CCOTTGAGGA CGGTGAGTGA CAGTTGAGTA ATAGAGAAAG CTGATATAAG ACTATGAGAAAG CTAGTATAAG ACTATGAGAGA ACTATGAGAGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTATGT TCTATCTTCC CTCCAACACA TAGTACAGTA TAGTACAGTG TAGACATGTA TAGAAAAGGA ACAGGAGTTA ACAGGAGTTA	ACCIONALE THORITANACTO ACCAGACTOC ACCAGACTOC GARCACATTA GEGARACACA GEGATITTAT GTGAGACACA GCAGTACAGA GCAGTACAGA CTAAGAGAC TCATTIATAA GCAGAACACA TCATTIATAA GCAGAACACA TCATTIATAA GCAGAACACA TCATCGCAGA AAAACCTGAG GAGAAACCTCA CATAAACTCA CATAAACTCA CATAAACTCA CATCAAACTCA CATCAACTCA CATAAACTCA CATAAACTCA CATCAACTCA CATAAACTCA CATCAAACTCA CATCAACTCA CATCAACTCAACTCA CATCAACTCAACTCA CATCAACTCAACTCA CATCAACTCAACTCA CATCAACTCAACTCA CATCAACTCAACTCA CATCAACTCAACTCA CATCAACTCAACTCA CATCAACTCAACTCAACTCA CATCAACTCAACTCAACTCA CATCAACTCAACTCAACTCA CATCAACTCA	AGGANTITTA GGAATITTA TATGCATCA TATGCATCA AATGGGAAA GCAAGGAAGG TACACTGAT GACTATGCA GTATGACTTA AATAAGATG TACATTATAC TACATTATAC TACATATACA TACATATACA TACATATACA CTATGACTTA AATAAGATC GTGGTTGTAT ACTGGGACCC GGGAACCAGA AATATATTCT	TACGATICECE TAAGGATAGA TTCATGCTGC AGATGTCTT TAAAGCAGCA ACACTCGTG AGAGCTGGTC AGAGCTGGTC ACCATCTGGG ACCATCTGACTAAAAAT TTAAGATAAAAT TAGGGGAAACCC ATAGAGTGAAT TTAAGAGTGAT TTAAGA	TGGGGNAAGG AGATAATACT CAGTATTAAT TCCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGAATAA ACAAGATAAT GGATGCCCAG GGAGACATAA TATGGAAGGC TCTCGTGGGA TCTCGTGGGA TTTTGTTGAC TGTTCCCTAC TGTTCCCTAC TGTTCCCTAC TGTTCCCTAC TGTTCCCTAC TATTATAAAA CCTCATTTT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTAGC TTACTAGGTT TGAGAGGGTTT TTITCATCTG GCATTCCAAG GAAGTCATTA TCTGGCAACA AGGCTAGAAC CAATAGATAG CTTTTGTGAG TATGATAGA TACAGATAG TACAGATAG AGGCTAGAG AGGCTAGAG AGGCTAGAG AGGCTAGAG AGGCTAGAG AGGCTAGG AGGCTAGG AGGCTATCCT AGGCCATCG	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTAAMA AACGTCAAGGA ACCTTAAGGAG CCCTAATGAA CGTAACAGTA GAGTCTTGGG GATTAGGAG GATTAGGAGT GATTAGGAGT GATTAGGGCAT CCGGACTAGA TGGTTCACGC AGATGAACAGT AGAGGAACAG AAAAAGGGAACAG AAAAAGGGAACAG AAAAAAGGGAACAG AAAAAAGGGAACAG AAAAAAGGGAACAG	ACCIGACITY TETGCAGCA TANCTUTCA NATITIGANA TOCCAGGTAC GATANATUT GATGAGTAG ATTICAGGAG ATTICAGAAA ACCITICAT TIACCITICA GATANAGAT TIACCITICA GATANAGAT TIACCITICA GATANAGAT TICGANAGAT CTGGAGAGT CTGGAGAGT CTGCAGC GAGTIAATCA AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA	TEGROCTEGE TTGTAGCEAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG CCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAAT CCGAGGGGG TTGAGTGAT AAGCCCTTAG AAGCCCTTAG AAGCCCTTAG CCAAGGGGT AAGCCCTTAG AACCACTTAG AACCACTAG AACCACTTAG AACCACTTAGAATTA TTAAAGTAGCA TTAGAAATTA AAAGCAAACCA TTAGAAATTA AAAGCAACCAC	GAMTANTGA AGAGAGATTT TCTGTAMAG TACTAGGGGT CTAMAGGGATC CTTAMATGT TTCCATATAG ATOGAGGTAT CGTACANGG ATOGAGGTAT TATGGACCAC CCTCATGAA TATGGACCAC CCTCATGAA TCAGGACACGT TCAGGAGAGG TCAGGAGACAC AGGGCATAAA ATATATAAG TGGATGGATA TTTAGACCAC	1411 1511 1711 1811 1911 2011 2211 2211 2311 2411 2511 2611 2711 2911 3011 3111 3111
CCOTTGAGGA CGGTGAGTGA CAGTTTAGTT ATGGGGATAA ATAGAGAAAC CTGATATAAG ACTATGACAA CGAGAGTATAAA ATATTAGGAA ATATTAGGAA ATATTAGGA ATATTAGGA ATATTAGGA ATATTATCT CTCCAACACA TATTTACTTCC CTCCAACACA TAGTACATTGT TAACTTGTA TGAACATGT TGAACTTGTA ACAGGAGTTA ACAGGAGTTA ACAGTAAATTA ACAGGAGTTA A	ACCIONAGE TITGCTGAGA TITGCTGAGA TTGTAAAGTG ACCAGACTCC GAACACATTA GCGCATTTTAT GTGAGACAGG GCATTCCTGA AGTAGAGG TCATTAAAGGAG CTAAGGAG CTAAGGAGC TCATTATAAA GCAGAAGCAT TAGGTGAGT CCTGCAGTA CCTGCAGTA CCTGCAGTA CCTGCAGTA TGCCCCGAGA AAAAGCTGAG GAGCAAAGCG CATAAACTGA TTAAAGGAG TTAAAGGAG TTAAAGGAG TTAAAGGAG TTAAAGGAG TTAAAGGAG TTAAAGGAG TTAAAGGAG TTAAAGGGC TTAAAGGAG TTAAAGGAG TTAAAGGGC TTAAAGGGC TTAAAGGAG TTAAAGGGC TTAAAGGC TTAAAGGGC TTAAAGGAC TTAAAGAC TTAAAGGAC TTAAAGCAC TTAAAGGAC TTAAAGGAC TTAAAGGAC TTAAAGGAC TT	AGGITMETT GGATTTTCATT TATGCATCA CTCTTTGATC AATGGGGAAA GCAAAGAAGG TACCACTGAT GACTATTGCA CTATGCACTTA TAGCACTTA TAGCACTTA TAGCACTTA TAGCACTTA TAGCACTTA TAGCACTTA TAGCATTATACA TTTGGAACTC CTGGTGTGTAT AGTGAGGACC GCGGAACCAGA AAATTATTCT AAAGAGAATT	TACGATUCCE TAAGGATAGA TTCATGCTGC TAAAGCAGCA ACATTGATAG ACACTGGTG ACGATGTATTA ATTATCTGCG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG TTTAAATAT TGGGAGAACCC ATAGGAGAACCC ATAGGAGAACC ATAGGTGAT TTAAGGTCATT TAGGTCATT	TEGGONANGE AGATANTACT CAGTATTAAT TECCTTTAGE CAITCAGACA TEAGGCCAGE ACAGGATAA CAAAGACTCA ATCAGATAAT ATCAGATAAT ATAGGAGAG CAGAGTCTGA TCTGGTGGGA CAGAGTACTA AAACAGTAGT TTTTGTTGAC TGTGCCCTAC TGTGTTGTGAC TATTATAAAA CCTGGTTTGTG	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTATTGAGTT TGAGAGGTTT TTTTCATCTG GCTTGCAAG GCAGTCATTA TCTGGCAACA AGGCTAGACA AGGCTAGACA ATCCTGAATG CTITTGTGAG TATGATATAC TACAGCAGGT AGGGCCAGGG ATGCCTTGTT AGGGCCAGGG ATGCCTTGTT AGGGCCATGG AGAGTTTCCT ACAATAGTCA AAGTAGGGCCA	ATANGGATGA GCANTGTGAT ACAGTTTCG ACACTACGG CAATATATGT ACGTCAGGG ACCTAAGGAG ACCTAAGGAG CCGTAATGAA CTGAACAGTT GATCATGG TITTAGTATG ATTAGGAG GTTTGGGG TTTTAGTATG ACTTAGGAG TTTTAGTATG ACTTAGGAG TTTGGGAG TTTTAGTATG ACTTAGGAG TGGTCAGCAT AGAGGAACAG AAAAAGGGAG AAAAAGGGAG AAAAAGGGAA AAAAAGGGAAAAA	ACCTGGCTT TCTGCAGGCA TAAACTCTCA AATTTTGAA TGGCAGGTAC GATAAATGT GATGAGTAGG ATTCGAGAA ATTGGACAAT TGACAGTAGG CAGTACAAA TGATAATGTG TTACCTTGCA GATAATGTGA TTACCTTGCA GATAATGGAA TTGGAGAGT TTGGAGAGT TTGGAGAGT TTGGAGAGT TTGCCCAGC GAGTTAATCA AGATCAACAA TAAATGGGAT TAAATGGGAG	TEGROCTEGE TTGTGTCATA TTTAMACAGT TTGTCTAGGG CCACTAGGTA COTTCTCT TTTTGTATGC AMATGCAT TGGTGAGGGG AAAGAGGGG TAGACGTGTA AACCCTTAG AAACTGCAT ACCATGAAT ACCATGAAT ACCATGAAT ACCATGAAT ACCATGAAT ACCATGAAT ACCATGAAT ACCATGAAT TAGAAGTAC CTTAGAATTA CAAGGCAGT ATTCAAGTGC TTAGAAATTA TAAGGAAGGA TAGAAATTA TAAGGAAGG	GAMTANTGA AGATAMTAAG AAAGGAGTTT TCTGTAMAAG TACTAGGGGT CTATAMTGT TTCCATATTGG ATGGAGGTAT CGTACCAAGA AGTTCAMATG TATGGACCAC CCTCTATGAMTG TATGGACCAC TCAGGACCAG TCAGGACCAG TCAGGACAGT TCAGGAGAGG TCAGGACAGT TCAGGAGAGG TATGGATACA ATATATAAAGG AGGATGGATA ATATATAAAGG AAAGAGCTAC AAAAGAGCTAC	1411 1511 1711 1811 1911 2011 2111 2211 2511 2611 2611 2911 3011 3111 3311 3411
CIOTEGAGGA CGGTGAGTGA AGGGGATAA ATAGAGAAAG CGAAGTATAAG ACTATGACAA ACTAGCAGAGA ACTATGAGCAA ACTATGAGGAAAG ACTATATAGGA ACTATTAGGA ACAATTATGT TOTATCTTCC TOCAACACA TATCTGACTA TAGGACAGA ACAATTATTAGGA TAGACAGTA TAGAAAAGGA ACAGGAGTTA ACAGGAGTA ACAGGAGTA ACAGGAGTA ACAGGAGAGTA ACAGGAGAGTA ACAGGAGTA ACAGGAGTA ACAGGAGAGTA ACAGGAGAGTA ACAGGAGTA ACAGGAGAGTA ACAGGAGAGTA ACAGGAGAGAGAGA ACAGGAGAGAGAGAGA ACAGGAGAGAGA	ACCIONALE THORTHANA ACCIONAL AL ACCIONAL AL ACCIONAL AL AL ACCIONAL AL AL ACCIONAL AL AL ACCIONAL AL A	AGGANTITTA GGAATTITA TATGCATTA AATGGGAAA GCAAGGAGG TAACACTGAT GACTATTGCA TACCACTGAT GACTATTGCA TACCACTGAT TACCACTGAT TACCACTGAA TACCACTGAT AAATAAGAT TACGATGTAT TACGACTGA TACGACCA GGGAACCAG AAATTATTCT AAAGGGACT AAAGGGACT AAAGGGACT AAAGGGACT AAAGGGACTGAA AAAGGGACTGAA	TACGATUCCC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAGGAGCA CAATTGATAG AGAGCTGGTC ACCATCTGGG ATTGTTCTGT TTAAGCTATTA GGGAGAACCC ATAGGAGGAT TTAAGGTCATT TAAGGATGAT TTAAGATTAT GAGAGAATTTA GATAGATTAT	TGGGGNANGG CAGTATTANT TCCTTTAGACA CATTCAGACA TGAGGCAGG ACAGAGATAN CANAGATAN TATGGAAGGC GGAGACATAN TATGGAAGG CAGAGTCTGA TCTGGTGGGA CAGAGTCTGA TTTGTTGGACAT TATTGTTGACT TGGCCTACT TGTCCCTAC TGGCATTTA TATTATANA AACAGAGA AACAGAGA CTTATTATA	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTOG TTACTGAGTT TTATTGATG GCTTTGCAAG GCATCATTA AGGCTAGAAC GAAATGGTA ATCCTGAATG CTTTTGTAAG CTTTTGTAAG TATGATATAC TACAGCACGT AGGCTCAGGA AGGCTCAGGA AGGCTCAGGA AGGCTCAGGA AGGCTCAGGA AGGCTCAGGA AGGCTCAGGA AGGCTCAGA AGGCTCAGA AGGCTCAGA AGGCTCAGA AGGTCAGA AGGTTCACA AGGTAGGGCC	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTAAGGAG CCGCTAATGAA CTGAACAGTA GAGTCTTGGG TTTTAGTATGAGG GTTTTAGTATGAGGAG GTGTGGGCAT GGACAGGTAGAA GGTTGGGCAT GGACTAGA TGGTTCGGCAT AGAGGAACAG AAAAAGGGAG AAAAAGGGAG AAAAAGGGAC AAAAAGGGAC AAAAAGGGAC CTTTCATTATA	ACCIGACITY TETGCAGCA TAMETETCA MATHTGANA TOCCAGGTAC GATAMATET GATGAGTAGG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG TAGATACAAA GACTITIGAT TAGATAGTAG GATAMAGAT TTAGGAAAGAT TTGGAAAGAT TTGAATAGTAGAT AGATTGAGAAG AGATAGGAGAG AGATAGGAGAG AATTGGAAA	TEGAGECTEGE TTGTAGECTAGT TTGTATCAGGG GCACTAGGTA TTTTGTATCAGGG GCACTAGGTA TTTTGTATCG AAAATGCAAT GCAAGGGCGC AAAAGAGGGGG TTGAGTTAG CAAAGATGCAAT AGCATGGATA AGCATGGATA AGCATGGATA ATTCAAAT AGCATGGATG ATTCAAGTGC TTAGAAATTA	GAMTANTGA AGAGAGTTT TCTGTAMAG TACTAGGGGT CTAMAGGGTTC CTTAMATGT TCCTACAAG ATGGAGGTAT CGTACCAAG AGGTATATGAGACAC CCTCTAGAATGA TATGGACCAC CCTCTAGAA TCAGCACCAC TCAGGACCAC TCAGGACCAC TCAGGACCAC TCTGATACC AGGGCATAAA ATATATAAGG TGGATGGATA TTTAGACCAC AAACAGCTAC AACCACCATCA	1411 1511 1711 1911 2011 22111 22111 2311 2411 2511 2611 2711 2911 3011 3111 3211 3311 3311
CCOTTGAGGA CGGTGAGGA CAGTTGAGTA ATAGGGAAAA ATAGAGAAAC CGAAGTATAAG ACTATGACAA CGAAGTATAAG ATATTAAGGA ATATTAAGGA ATATTAGGA TATATTTC CGCCAACACA TAGTACAGTG TAAATA TGAAAAAGGA ACAGGAGTTA TGAAAAAGGA ACAGGAGTTA TGACTACACT TGACTACACT TGACTTAGTA	ACCIONADA TITGCTGAGA ACCAGACTCC GAACACATTA GGGATTTTAT GTGAGACAGG ACATANGGAG GCAGTCCTGA AGTAGAAGTG CTAAGAGAC TCATTTATAA GCAGAACCAT TAGGTGAGTA CCTGGCAGTA TAGCTGAGTA CCTGGCAGTA TAGCTGAGTA CATAAACTGA GAGCAAACCA TAAAACCTGAG GATCAACCA TAAAGAGCT	AGGANTITIC TGTTTGCATT TATGCATCA CTCTTTGATC AATGGGAAA GCAAGGAAGG TACCACTGAT GACTATGCA TACCACTGAT TACCACTGAGACCC GGGAACCACA AAATTATTCT AAAGGAATT ACACCTGTGA ACCCTGTGAG	TAGGATHCECE TAAGGATAGA TTCATGCTGC AGATGTCTT TAAAGCAGCA AGATGTCTGTG AGAGCTGGTC AGAGCTGGTC ACCATCTGGG GTCTGACTGC ACCATCTGGG GTCTGACTGC ACCATCTGTGGG GTCTGACTGC ACTACTGCTAAA ATTTATATAT GGGGAGAACCC ATAGATGTATTAAATAT TAAGGTGATT TAAGGTGATT GAGAAATTAA GATAGATTTA	TGGGGMAAGG TGATTAAT TCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGGACATAA ATCAGATAAT GGATGCCCAG GGAGACATAA TATGGAAGG TCTCGTGGGA TCTCGTGGGA TCTCGTGGGA TTTTGTTGAC TGTTCCCTAC TGTTCCCTAC TGTTCCCTAC TATTATAAAA CCTGATTGTG AAACAGAAGA TATTATAAAAA CCTGATTGTG AAACAGAAGA CTTATTATAAAA CCTGATTGTG AAACAGAAGA CTTATTATAAAA	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTACTAGGTT TGAGAGGTTT TTITCATCTI GCATTGCAAG GCATTGCAAG TCTGGCAACA AGGCTAGAAC CATATGATG CATTTGTGAG TACGACGT AGGGTCAGGG ATGCCTTGTA AGGGTCAGGG AGGGTTCCT AGAGTAGCA AGATAGTCA AGATAGTCA AGATAGTCA AGATAGTCA AGATAGTCA AGTTGTTAAA	ATANGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTANAA AACGTCAAGAG ACCTAAGGAG ACCTAAGGAG CTGAACAGTA GAGTCTTGGG GATTAGGAG GATTAGGAG GATTAGGAG GATTAGGAG GATTAGGAG GATTAGGAG GATTAGGAG GATTAGGAG CTGACTTAGTAG AACATGAAA AACATGAAAG CCTAATTAC CCTGATTAGA	ACCIGACITY REGICAR TANACTUCA ANTITIGANA TOGCAGGIAC GATANATUT GATAGGIAG ATTOGACAT ACGATAGA ACGATACAAA ACGATACAAA ACACTITGGA TIACCITIGCA GATAATGGATG CTGGAAAGAT TTGGAAAGAT TTGGAAAGAT CTGGAAAGAT ACTAATGGATA AGATCAACAA AGATCAACAA TAAATGTGAT AATTGGAAAA TAAATGTGAT AATTGAACAA ACATCAACAA TAAATGTGAT AATTGAACAA TAAATGTGAT TAATGAACAA TAAATGTGAT TAAATGTGAT AATTTGAACAAA ATTTGAACAAA TTTGAACAAA TTTGAACAAAA TTTGAACAAA TTTGAACAAAA TTTGAACAAAA TTTGAACAAAA TTTGAACAAAA TTTGAACAAAA TTTGAACAAAAAAAAAA	TEGROCTEGE TTGTAGECTAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA CGTTCTCTT TTTTGTATGG AAAATGCAAT CGGAGGGGG TTGATGTGAT AAAATGCAT AAGCCCTTAG CCAAGGGGG TTGATGTTAC AATCTGAAT CAAGGCCTGT CAAGGCCTGT TATCAAGTTGC TTAGAAATTA TAAGGAAGA AAAAGAAGAA TTAAAGTTCC TTAAAGTTCC TTAAAGTTCC TTAAAGTTCC TTAAAGTTCC TTAAAGTTCC TTAAAGTTCC	GAMTANTGA AGAGAGTTT TCTGTANAAG TACTAGGGGT CTTAMATGT TCCATTATGG ATOGAGGTAT CGTACANGG ATOGAGGTAT TCAGGACCAC CCTCATGAA TCAGGACCAC CCTCATGAA TCAGGACACAC TCAGGAGAGG TCAGGACACAC TCAGGAGACAC ATOTATATAGACAC TGGATGGATA TTTAGACAG AAAGAGCTACA AAACAGCATAA	1411 1511 1711 1811 1911 2011 2111 2211 2311 2511 2611 2711 3011 3111 3211 3311 3511 3511
CCOTTGAGGA CGGTGAGGA CAGTTGAGTA ATAGAGAAAG CTGATATAAG ACTATGACAAA CGAAGTATAA CTAGCAGAGA ATATTAAGG ATATTAAGG ATATTAAGG ATATTAAGGA ATATTAAGGA ATATTATGT TCTATCTTCC GTCCAACCA TAGTACAGTA TAGTACAGTG TAACTTGTA TAGAAGAGTTA ACAGAAGTTA ACAGAAGTTA ACAGAAGTTA ACAGAAGTTA CAGTTAATC TGACTTGTCC TGTACAGTCC TGTACAGTCC TGTACAGTCC TGTACAGAGA CAGTTACAGAC CAGTTACACAC CAGTTACACACAC CAGTTACACACAC CAGTTACACACAC CAGTTACACACAC CAGTTACACACAC CAGTTACACACAC CAGTTACACACACAC CAGTTACACACACAC CAGTTACACACAC CAGTTACACACAC CAGTTACACACACACACAC CAGTTACACACACACACACACACACACACACACACACACA	ACCIONALE THORITANACTO ACCAGACTOC ACCAGACTOC GANCACATTA GGGATTTTAT GTGGAGACAGG ACATANAGGG ACATANAGGG CTAAGAGAC TCATTTATAA GCAGAACACT TAGTTGGTAG ACTGGCAGA AAAAGCTGAG AAAAGCTGAG CATAAAGAG CATAAAGAG CATAAACTGA GAGCAAACCA TTAAAGAGCT TAAAGAGCT TAAAGAGCT AAAGCTCACG AACCTCCTACG AAGCTCACAG AAAGCTCACAG AAAGCTCACAG CATCACCCCACAG AAAACCTCACCACAG AAAGCTCACAG CATCACCCCCACAG AAAGCTCACAG CCTCTCCTTTC	AGGANTITIC TGTTTGCATT TATGCATCA TCTCTTTGATC AATGGGAAA GCAAGGAAGG TACCACTGAT CACTATTGCA TACCACTGAT AATAAGACA CCTGGAGCCC CGGGAACCAGA AAATAATTCT AAAGAGAATT ACACCTGTGA TCCACCTGTGAGGGA TCAGTTCCAC TTCCATTCATTCTT TACCACTGAGGGA TCAGTTCCAC TTCCATTCATTCATTCATTCATTCATTCATTCA	TAGGATHCECE TAAGGATAGA TTCATGCTGE AGATGTCTT TAAAGCAGCA AATAGATAG AGAGCTGGTC AGAGCTGGTC AGAGCTGGTC ACCATCTGGG GTCTGACTGC ACCATCTGGG GTCTGACTGC ACCATCTGGG GTCTGACTGC ACTAGTGCCTA TIAGCTATIA ATTCTGT GCTCAGTAAA TCTTTAAATAT TAGGATAGA TAGAGTGAT GAGAGATACT GAGAAATTTA GAGAAATTTA GAGAAATTTA GAGAAATTTA GAGAAATTTA GAGAAATTTA GAGAAATTTA GGCACAAGAT AGCCCACAGAT	TGGGGNANGG CAGTATTATT CCCTTTAGG CATTCAGACA TGAGGCAGG AGAGAGATAA CAAAGATCA ATCAGATAAT ATTGAAGGCCAG GGAGACATAA TATGGAAGGC CAGAGTCGA TCTGGTGGA TCTGGTGGA TTTGTTGAC TGTGCCTTAC TGTACATTAAAT TATTATAAAA CCTGATTGTG AAACAGAAGA CTTATTTTAAAA CCTGATTGTG AAACAGAAGA TGATTCATAACT TGATTAAACT TGATTAAACT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTACC TTATTATTACCT GCATTCCAAG GCTTTCCAAG GCATTCCAAG GAAGTCATTA TCTGGCAACA AGGCTAGAAG CTTTTGTGAG CATTTGTGAG CATTTGTGAG AGGCTCAGG AGGCTCAGG AGGCTCCTTCTT AGGCCATCG AGACTTTCCT ACAATAGTCA AAGTAGGCC TACTGTTAAA CTATTTCTGT GCACATCAAG CTATTCTTTA	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTAAMA AACGTGAGGT ACCTAAGGAG CCCTAATGAA CGTGAGGT GATCTTGGG GATCTAGGAG GATCATGGG GATCAGGAA GATCAGGAA GATCAGGAA GATCAGGAA TGGTTCACGC AGAGGAACAG AAAAAGGGAA AAAAAGGGAA CCCGACTAGAA AACATGAAAG CCTAATTATA AACATGAAAG CCTAATTTACAAG CCTAATTTACAAG AATATTCAAG CATATTTCAAG CATATTTCAAG CATATTTCAAG CATATTTCAAG CATATTTCAAG CATATTTCAAG CATATTCAAG CATATTTCAAG CATATTCAAG CATATTTCAAG CATATTCAAG CATATTCA	ACCTGGCTT TETGCAGCA TAAACTCTCA NATITIGAAA TGCCAGGTAC GATAAAATGT GATGAGTAGG ATTTCAGGAG ATTGGACAAT TTACCTTGCA GACTACTGGT TTACCTTGCA GATAATGGAT CTGGAAGAGT CTGGAAGAGT CTGTGCCAGC GAGTTAATCA AGATCAACAA AGATCAACAA AGATCACAG AGATCACAA TAAATGGAT GAGTAGGAG ATTTGACAAA TAAATGGAT GAGTAGGAG ATTTGACAAA TAAATGGAT GAGTAGGAG ATTTGACAAA	TEGROCTEGE TTGAGCEAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG CCACTAGGTA CGTTCTCTCT TTTTGTTATGG AAAATGCAAT CCGAGGGGG TTGAGTGAT AAGCCCTTAG CCAAGGTAC AAATGCAAT AAGCCCTTAG CCAAGGATA CAAGGCCTGT CCTAGCAAGT AATCAAATTAC AATCAAATTA CAAGGCTGT TAAGAATTA AAAAAAATAC TTAAGAATTA AAAAAAAAAA	GAMTANTGA AGAGAGTTT TCTGTAMAG TACTAGGGGT CTAMGGGGTC CTTAMATGT TCCATATAGG ATGGAGGTAT CGTACANGG TATGGACGAC TATGGACCAC TCAGGACACG TATGGACCAC TCAGGACACG TCAGGACACG TCAGGACACG TCAGGACACG TCAGGACACG ATTATATAGG AAGAGCTACA AACAGCTACA AGAGCTACA AGAGCTACA AGAGCTACA CCACTACACAC CCCTCACTGAA	1411 1511 1711 1911 2011 2112 2211 2311 2513 2611 2711 2611 2911 3011 3111 3411 3411 3411 3611
CCOTTGAGGA CGGTGAGGA CAGTTTAGTT ATGGGGATAA ATAGGGAAAG CTGATATAAG ACTATGACAA CGAGGTGAGA ATATTAAGGA ATATTAGGAA ATATTAGGAA ATATTAGGA ATATTAGGA ATATTATGT TGATTATCTTCC CTCCAACACA TACTTGACTA TACTTGACTA TGAAAAGGA ACAGGAGTTA ACAGGAGTTA ACAGGAGTTA TGAAAAGTA TGAAAAGTA TGAAAAGTA TGAAAAGTA TGATTACTCC TCTACAGTCA CTTACAGTCA CTTACAGTCA CATTATACAGTA CAGTTAAAGGA AGAGGTTA CAGTTACAGTCA CATTATACAGTA CAGTTACAGTCA CATTATACAGTA CAGTTACAGTA	ACCIMANDA TITGTIGAGA ACCAGACTICA GARCACATTA GGCATITTAT GTEAGACAGG ACATANGGAG GCAGTCCTGA AGTAGAGAG CTAAGAGAA CCTGGCAGTA CCTGGCAGTA CCTGGCAGTA ACAGCTGAG GAGCANAGCG CATAAACTGA GAGCANAGCG CATAAACTGA GAGCANAGCG CATAAACTGA GAGCANAGCG CATAAACTGA GAGCANAGCG CATAAACTGA AAGGCTCACCA GACCTCCTTG CTCTTTTATC CTCTTTTTATC	AGGITITTA GGAATTTTA TATGCATCA CTCTTTGATC AATGGGAAA GCAAAGAAGG TACCACTGAT GACTATGCA CTATGCA CTATGCA CTATGCA CTATGCA CTATGCACTTA AAATAAGACTC CTGGAACTC CTGGAACTC CCTGGAGCCT CCTGGAGCCT CCTGGAGCCT CCTGGAGCCT CCTGGAGCCT CCTGTGAGACT AAAGAAATTATCCT AAAGAAATTATCCT AAAGAGACT AAAGAGACT ACACCTGTGA CTCTAGAGGA TCAGTTCCAC TTCATTATATT TCATTTTATTTCT TCATTTTATTTCT TAGAGGACT TCAGTTCCAC TTTCATTTTATTCT TAGAGGACT TCAGTTCCAC TTTCATTTTATTTCATTTTATTCT ACACCTTCCAC TTTCATTTTATTTCATTTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCATTTTATTCA	TACGATTCCC TAAGGATAGA TTCATGCTGC AGATGTCTT TAAAGCAGCA CAATTGATAG ACACTCGTG ACACTCGTG ACTCTCGTG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGT TAACGTATAA ATTGTCTGT GCTCAGTAAA TCTTAATATT TAGGTCAT TAAGATTTAATAT TAGGTCAT TAAGATTTAATAT GAGAAATTTA GAGAAATTTA GAGAAATTTA GAGAAATTTA CCCTAAACAA TCCCTAAACAA TTACGTTATA	TGGGGNANGG AGATANATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAGGATAA CANAGACTCA ATCAGATAAT TGCATTCTGGGA CAGAGTCTGA TCTGTGGGA CAGAGTCTGA ANACAGTAGT TTTTGTTGAC TGTCCCTAC TTTTTTTAAAA CCTATTTAAAA CCTATTAAAC TCAATTAACT TCAATTAACT TCAATTAACT TCAATTAACT TCAATTAACT TCAATTAACT TCAATTAACT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTATTGAGTT TGAGAGGTTT TTATCATCTG GCTTTGCAAG GCATTGCAAG AGGCTAGACA AGGCTAGACA ATCCTGAATG CTATTGTGAG ATCCTGAATA TATCATATATAC TATCATATATAC TACAGCACGT AGGGCTAGGG AGGCTTTCCT ACAGCACGT ACAGCATGGC ACATAGCA AAGTAGTCA AAGTAGGGC TACTGTTAAA CTATTTCAA CAACTATTAA CAACTATTAA	ATANGGATGA GCANTGTGAT ACACTATCG ACACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTTAGGAG CCGTAATGAA CTGAACAGTA GAGTGTGGG TTTTAGTATG GATGAGGAG TGTTGGGCAT CCGGACTAGA TGGTTCACTC AGAGGACAGA TGGTTCACTC AGAGGACAGA TGGTTCACTC AGAGGACAGA TGGTTCACTC AGAGGACAGA TGGTTCACTC AGAGGACAGA ACAATGAAG CCTAATTTAC CCTGACTGAT AATATTCAAG CATAGCATTT	ACCTGGCTT TCTGCAGGCA TAAACTCTCA AATTTTGAA TGGCAGGTAC GATAAAATGT GATGAGTAGG ATTCGAGAA ATGGACAAT TGACATGACA	TEGAGCETEG TTGTGTCAGG TTGGTTCTTA TTTAMACAGT TTGTTCAGGG CCACTAGGTA COTTCTCT TTTTGTATGG AMATGCAT TTGTATGG AMATGCAT TGGTAGGAG AMAGGGGG TTGATGTAT ACCATGCATA ACCATGCATA ACCATGCATA ACCATGCATA ACCATGCATA ACCATGCATA ATTCAAGTGC TTAGAAATTA TAAGGAAGA AMAAGAGAT TTAAGGTTC TTTAGAATTA ACATGCATA AAAGAGATA AAAAGAGAT AAAAGAGAT AAAAGAGAT AAAGATTCATTCAATGC TTTAGAATTA AAAGAATTA AAAGAATTA AAAGAATTA AAAGAATTA AACGATTCAAT	GAMTANTGA AGAGANTTAG AAAGGAGTTT TCTGTAMAGT TACTAGGGGT CTATAGGGGTTT TCCACCAGA AGTGAGTAT TCGACCAGA AGTTCAMATG TATGGACGATA TCAGCACCAC AGTCAMAAAT TCAGCACCAC TCAGGAGAGG TTCAGGAGAGG TTCAGGAGAGG TCAGGATAA AGTATATAAAGG TGGATGGATA AGTAGACACAC AAACAGCAT CCACCACCAC AGGACTTGTA AACACCACCAC CCTCACCACAC AGGACCTGTA AGGACCTTGTA AGGACCTAGA AGGACCTTGAA AGGAACCTAGA	1411 1511 1711 1811 2011 2211 2211 2211 2311 2411 2511 2711 2911 3011 3111 3111 3511 3511 3713 3811
CCOTTGAGGA CGGTGAGGA CAGTTTAGTT ATGGGGATAA ATAGGGAAAG CTGATATAAG ACTATGACAA CGAGGTGAGA ATATTAAGGA ATATTAGGAA ATATTAGGAA ATATTAGGA ATATTAGGA ATATTATGT TGATTATCTTCC CTCCAACACA TACTTGACTA TACTTGACTA TGAAAAGGA ACAGGAGTTA ACAGGAGTTA ACAGGAGTTA TGAAAAGTA TGAAAAGTA TGAAAAGTA TGAAAAGTA TGATTACTCC TCTACAGTCA CTTACAGTCA CTTACAGTCA CATTATACAGTA CAGTTAAAGGA AGAGGTTA CAGTTACAGTCA CATTATACAGTA CAGTTACAGTCA CATTATACAGTA CAGTTACAGTA	ACCIMANDA TITGTIGAGA ACCAGACTICA GARCACATTA GGCATITTAT GTEAGACAGG ACATANGGAG GCAGTCCTGA AGTAGAGAG CTAAGAGAA CCTGGCAGTA CCTGGCAGTA CCTGGCAGTA ACAGCTGAG GAGCANAGCG CATAAACTGA GAGCANAGCG CATAAACTGA GAGCANAGCG CATAAACTGA GAGCANAGCG CATAAACTGA GAGCANAGCG CATAAACTGA AAGGCTCACCA GACCTCCTTG CTCTTTTATC CTCTTTTTATC	AGGITITTA GGAATTITTA GGAATTITAT TATGCATCA ATTGCATCA ATTGGGGAAA GCAAAGAAGG TACCACTGAT GACTATTGCA CTATGCACTTA TAGCACTTA TAGCACTTA AAATAAGACTC GTGGTGGTAT ACTGGACCTC GTGGAGCCT GCGGAACCAGA AAATTATTCT ACTGGAGCCT GGGAACCAGA AAATTATTCT AAAGAGATT ACACCTGTGA GTCTAGAGGA TTTGATTATACA TTTGATTTCATTATA	TACGATTCCC TAAGGATAGA TTCATGCTGC AGATGTCTT TAAAGCAGCA CAATTGATAG ACACTCGTG ACACTCGTG ACTCTCGTG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGT TAACGTATAA ATTGTCTGT GCTCAGTAAA TCTTAATATT TAGGTCAT TAAGATTTAATAT TAGGTCAT TAAGATTTAATAT GAGAAATTTA GAGAAATTTA GAGAAATTTA GAGAAATTTA CCCTAAACAA TCCCTAAACAA TTACGTTATA	TGGGGNANGG AGATANATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAGGATAA CANAGACTCA ATCAGATAAT TGCATTCTGGGA CAGAGTCTGA TCTGTGGGA CAGAGTCTGA ANACAGTAGT TTTTGTTGAC TGTCCCTAC TTTTTTTAAAA CCTATTTAAAA CCTATTAAAC TCAATTAACT TCAATTAACT TCAATTAACT TCAATTAACT TCAATTAACT TCAATTAACT TCAATTAACT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTATTGAGTT TGAGAGGTTT TTATCATCTG GCTTTGCAAG GCATTGCAAG AGGCTAGACA AGGCTAGACA ATCCTGAATG CTATTGTGAG ATCCTGAATA TATCATATATAC TATCATATATAC TACAGCACGT AGGGCTAGGG AGGCTTTCCT ACAGCACGT ACAGCATGGC ACATAGCA AAGTAGTCA AAGTAGGGC TACTGTTAAA CTATTTCAA CAACTATTAA CAACTATTAA	ATANGGATGA GCANTGTGAT ACACTATCG ACACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTTAGGAG CCGTAATGAA CTGAACAGTA GAGTGTGGG TTTTAGTATG GATGAGGAG TGTTGGGCAT CCGGACTAGA TGGTTCACTC AGAGGACAGA TGGTTCACTC AGAGGACAGA TGGTTCACTC AGAGGACAGA TGGTTCACTC AGAGGACAGA TGGTTCACTC AGAGGACAGA ACAATGAAG CCTAATTTAC CCTGACTGAT AATATTCAAG CATAGCATTT	ACCTGGCTT TCTGCAGGCA TAAACTCTCA AATTTTGAA TGGCAGGTAC GATAAAATGT GATGAGTAGG ATTCGAGAA ATGGACAAT TGACATGACA	TEGAGCETEG TTGTGTCAGG TTGGTTCTTA TTTAMACAGT TTGTTCAGGG CCACTAGGTA COTTCTCT TTTTGTATGG AMATGCAT TTGTATGG AMATGCAT TGGTAGGAG AMAGGGGG TTGATGTAT ACCATGCATA ACCATGCATA ACCATGCATA ACCATGCATA ACCATGCATA ACCATGCATA ATTCAAGTGC TTAGAAATTA TAAGGAAGA AMAAGAGAT TTAAGGTTC TTTAGAATTA ACATGCATA AAAGAGATA AAAAGAGAT AAAAGAGAT AAAAGAGAT AAAGATTCATTCAATGC TTTAGAATTA AAAGAATTA AAAGAATTA AAAGAATTA AAAGAATTA AACGATTCAAT	GAMTANTGA AGAGANTTAG AAAGGAGTTT TCTGTAMAGT TACTAGGGGT CTATAGGGGTTT TCCACCAGA AGTGAGGTAT TCGACCAGA AGTTCAMATG TATGGACGACA CCTCTATGAA TCAGCACCAC TCAGCACCAC TCAGGAGAGG TTCAGGAGAGG TTCAGGAGAGG TCAGGAGAGGAT AGAGCACACA AGAGCTTCA AACACCACCAC AGAGCTTGTG TCAGCACCAC AAGACCTTGT CCACCACCACAC AGAGCTTGTA ACACCACCAC AGAGCTTGTA AGAGCTTGTA AGAGCTTGTA AGAGCTTGTA AGAGCTTGTA AGAGCTTGTA AGAGCTTGTA AGAACCTAGA AGAACCTTGAA AGAACCTTGAAA AGAACCTTGAAA AGAACCTTGAAA AGAACCTTAAA AGAACCTTAAA AGAACCTTAAA AGAACCTTAAA AGAACCTTAAA AGAACCTTAAA AGAACCTTAAA AGAACCTT	1411 1511 1711 1811 1911 2011 2211 2311 2411 2511 2611 2611 3011 3111 3311 3411 3511 3611 3711 3611 3711 3611
CICTTGAGGA CGGTGAGGA CAGTTGAGTA ATAGAGAAAC CTGATATAAG ACTATCACAA CTAGCAGAGA ATATTAAGG ATATTAAGG ATATTAGGA TATTTTC CTCCAACCA TATTTGACTA TAGTACAGTG TAAACTTGTA ACAGGAGTTA ACAGGAGTTA ACAGGAGTTA TIGTCTCACT TIGCTCACT TIGCTCACT TIGCTCACT ACATTACAGA AGATGATTA AGGGGGTCTT AGGGGGTCTT AGGGGGGTCTT AGGGGGGCCTT TTGCTAGGGA	ACCIMANDA AGGGGCAGG TITGCIGAGA ACCAGACTCC GAACACATTA GGGATITTAT GTGAGACAGG ACATANGGA CCTAAGAGAC TCATITATAA GCAGAGACCATTA TAGGTGGCAGTA TAGGTGGGTA TOCCCGAGA AAAAGCTGAG GAGCAAAGCT TAAAGAGCT TAAAGAGCT TAAAGAGCT TAAAGAGCT CATTAACAGC GATCCTCCTAG GCTCTCCTTA CATTAAAAG GCACACATT CATTAAAAG GCACACACATT CATTAAAAG GCACACACATG CATTAAAAG	AGAITMETT GGAATTTTC GGTTTGCATT PATGCATCAA CTCTTTGATC AATGGGAAA GCAAGGAAGG TACCACTGAT GACTATGCA CTATTGCA CTTGCAGCCC CGGGACCAGA AAATTATTCT AAAGAGAATT ACACCTGTGA CTCTTGCAGCC TTTCATTATT CGAGGTTGTTATT CGAGGTTGTTATT CGAGGTTGTTT TCCATGCTTTT TGGAGGTTGTTT	TAGGATHCECE TAAGGATAGA TTCATGCTGE AGATGTCTT TAAAGCAGCA ACACTCGTG AGAGCTGGTC AGAGCTGGTC ACCACTGGG GTCTGACTGC ACCACTGGG GTCTGACTGC ACCACTGGGA ACTACTGCGA ACTACTGCGA ACTACTGCGA ACTACTGCGA ACTACTGGG GTCTGACTGA TTAACTATTA TATACTATTA TAGGATACA TTAAGATTAA TATAGATGATTA ACTAGAGTACT AGACACAGAT CCCTAAACAA ACTGGGGAAC ATTAGGTATAA ACTGGGGAAC ACCACAGAT ACTGGGGAAC ACTGAGGAAC ACTGGGGAAC ACTGGGGAAC ACTGGGGAAC ACTGGGGAAC ACTGGGGAAC ACTGGGGAAC ACTGGGGAAC ACTGGAGGAAC ACTGGAGGAAC ACTGGAGAC ACTGGAGAC ACTGGAGAC ACTGGAGAC ACTGGAGAC ACTGGAGAC ACTGGAGAC ACTGGAGAC ACTGAGAC ACTGGAGAC ACTGAGAC ACTGGAGAC ACTGAGAC ACTGGAGAC ACTGGAGAC ACTGAGAC ACTGAC ACTGAGAC ACTGAGAC ACTGAGAC ACTGAGAC ACTGAGAC ACTGAGAC ACTGAGAC ACTGAGAC ACTGAC ACT	TGGGGNAAGG AGATAATACT CAGTATTAAT TCCCTTTAGG CATTCAGACA TCAGGCCAGG AGAGAGAATAA ATAGATAAT ATAGAATAAT ATAGAATAAT ATAGAAGCCCAG GGAGACATAA ATAGAAGAG CAGAGCCCAG CAGAGCCTGAA TTTTGTTGAC TGTTGCCCTAC TGTTGCCCTAC TGTTGATTAAAA ACTAGTAGT AATTATAAAA CCTGATTGTG AAACAGAAGA TGTTTATATAAA TGTAATTATAAA TGTAATTATAAA TGTAATTATAAA TGTAATTATAAA TGTAATTATAAA TGTAATTAAAA TGTAATTAAAA TGTAATTAAAA TGCAATTAAAC TAAATTAACT TAAATTAAAC TTAAATTAAAC TTAAATTAAAC TTAAATTAAAC TTAAATTAACT TAAATTAAAC TTAAATTAAAC TTAAATTAACT TAAATTAACT TAAATTAACT TAAATTAACT TAAATTAACT TAAATTAACT TAAATTAAATT TAATTAAAAT TAATTAAATT TAATTAATT TAA	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTACC TTATTATTACC GCATTCCAAG GCATTCCAAG GCATTCCAAG GCATTCCAAG GCATTCCAAG TCTGGCAACA AGCCTAGAAG CTTTTGTGAG CATTTGTGAG ATCCTGATG AGGGTCAGGG AGGCTCACG AGGCTCACG AGGCTCACG AGGCTCATGC AGGTCAGGG CACTGTTAAA CAACTATTCAG GCAGATCAAG CAACTATTTA AAGCCATTTT GGATCCAGGC CACTATTTA	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTAAMA AACGTGAGGT ACCTAAGGAG CCCTAATGAA CGTGAGGT GATCTTGGG GATCTATGGG GATCATGGG GATCAGGAA GATCAGGAA GATCAGGAA GATCAGGAA TGGTTCAGGA TGGTTCACT AGAGGAACAG AAAAAGGGAA CCTAATTAAA ACATGAAAG CATAAGCATTT AATATCAGAG AATATCAGAG ATAGCATTT TATATCAGGA ATAGCATT TATATCAGGAG ATGGATGCAA ATGGATGCAA TGGATGCAAT	ACCTGACTT ACTGAGCA TANACTCTCA ANTITIGANA TOGCAGGTAC GATANATGT GATGAGTAGG ATTGAGCAGA ATTGAGCAGA ATGGACAAT AGATACTGG CAGTIACAAA GACTITGGT TTACCTIGCA GATAATGGAT TTGGANAGT TTGGANAGT TTGGANAGT TTGCCCAGC GAGTTAATCA GATAACGAA AAATGGAT GAGTAGGAG ATTTGACAAA CACTTGGACAC AGATCAACAA GATAGGAG ATTTGACACA TTAAGGGTG ACATTGAACT AAAAAAATGG ACATTGACCAT TTAAGGGTG TTTAGGACCAT TTTAGGACCAT	TEGROCTEGE TTGAGGCAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG CCACTAGGTA CGTTCTCTCT TTTTGTTATGG AAAATGCAAT CCGAGGCGCC AAAAGAGGGG TTGAGTGAT AAGCCCTTAG AAGCCCTTAG AAGCCCTTAG AAGCCCTTAG AATCGAAAT AACAAGTAC CCAAGCAGT TAAGAAGTA AAGCACTTTAAGAATTA AAGAAGAC TTAAGAAGT TAAGAAGAC ATGATTCAAATTA AAGAAGAC TTAAGAATTA AAGAAGAC ATGATTCAAT ACCTTCAAT ACCTATTAAG ACCTATTAAG CCACAGGATCT CCAAGGGCCA TCCAGGGCCAC	GAMTANTGA AGAGAGTTT TCTGTAMAG TACTAGGGGT CTTAMATGT TTCCATATGG ATOGAGGTAT CCTTAMATGT TCTACANGG ATOGAGGTAT TATGAGCACAT TATGAGCACAT TATGACCACAT TATGACCACAT TATGACACAC CCTCTATGAA TCAGCACACAT TCAGGAGACAC TCAGGAGACAC AGAGCATAAA ATATATAGCAG AAAAGAGCTAC AACAGCACAC CTCTACTAAA AGAAACCTACA AGAAACCTACA AGAAACCTACA AGAAACCTACA AGAAACCTACA AGAAACCTACA AGAAACCTACA AGAAACCTACA AGAAACCTACA AGAAACCACAC CTCTACTGAA AGAAACCTACA AGAAACCACAC AGTAAATTAGC AGTAAAATTAGC AGTAAATTAGC AGTAATTAGC AGTAAATTAGC AGTAAATTAGC AGTAAATTAGC AGTAAATTAGC AGTAAATTAGC AGTAAATTAGC AGTAAATTAGC AGTAAATTAGC AGTAAATTAGC AGTAATTAGC AGTAATTAGC AGTAAATTAGC AGTAGC AGTAG	1411 1511 1711 1911 2011 22111 22111 2311 2411 2511 2611 2711 2911 3011 3111 3211 3311 3311 3311 3311 3411 3711 3911
CICTURAGGA CGGTGAGGA CGGTGAGGAA ATAGAGAAAC CTGATATAAG ACTATGACAA CTAGGAGAA ATATAGGAAAC CTAGTATAAG ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTATGT ATATTATCT CTCTAACACA TAGTACATTA TAGTACATTGT TAACTTGTA ACAGTAAAT ACAGTAGTA TAGAAGTTA TGGATAACTTA TGGATTACTC CCTTAACTTA TGGATTACTC CTTACAGTC CTTACAGGC CTTACAGC CTTACAGGC CTTACAGC CTTACAGGC CTTACAGC	ACCIMANDA AGGGGCAGG TITGTIANAGTG ACCAGACTCC GANCACATTA GGCAITTTAT GTGAGACAGG ACATANGGAG GCAGTCCTGA AGTAGAGAG CTAAGAGAG CTAAGAGAG CCTGGCAGTA CCTGGCAGTA ACAGCTGAG GAGCANAGCG CATALACTCA GAGCAGAGGC CATALACTCA CATALACTCA CATCCTGCCA CATALACTCA CATCCTCCTTG CATCCTCCTTG CTCTTTTTATA GCACCAATT GAGAGGTGAG GCACCAATT GAGAGGTGAG GGACCAATT GAGAGGTGAG GGACCAATT GAGAGGTGAG GGACCAATT GAGAGGTGAG GGACCAATT	AGAITMETT GGAATTTTC TETTTGCATT TATGCATCA CTCTTTGATC AATGGGAAA GCAAAGAAGG TACCACTGAT GACTATGCA CTATGCACTTA TGGCTCTTTA AATGAACTC CTATGAACTC CTGGAACTC CTGGAACTC CCTGGAGCCT GCGAACCAGA AAATTATTCAT ACTAGAGACT TACAGAGACT TACAGAGCT TACAGCCT TACAGCC TACAGCCT TACACC TACAC	TAGGATHCECE TAAGGATAGA TICATGCTGC AGATGTCTT TAAAGCAGCA CAATTGATAG ACACTCGTG ACACTCGTG ACTCGTG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGT TAACGTATAAAAT TGTTAATATT GGGAGAACCC ATAGAGTGAT TAAGATTTAAGATTTAGGAAAATAT CGCAAGATTAGATATTAGATATTT GAGAAATTTAGGAAAACAA TATAGATATTAGACAAGAT CCCTAAACAAA TAGGATAATATTAGGCAATATAGAATATTAGGCAATATAA ACCACAAGAT CCCTAAACAAA TTAGGTGAAAC ACCACAAGAT CCCTAAACAAA TTAGGTGAAACAA TAGGTGGAAAC ACCACAACATTAAGATTAAGATATTAAGATATTAACATCATAGAACAA TTAGGTGGAAACAA TAGGTGGAAACAA TAGGATGGAAACAA TAGGTGGAAACAA TAGGTGGAAACAA TAGGTGGAAACAA TAGGTGGAAACAA TAGGTGGAAACAA TAGGATGAAACAA TAGGTGGAAACAA TAGGATGGAAACAA TAGGATGGAAACAA TAGGATGGAAACAA TAGGATGGAAACAA TAGGATGGAAACAA TAGGATGGAAACAA TAGGATGAAACAAA TAGGATGGAAACAAA TAGGATGGAAACAA TAGGATGGAAACAA TAGGATGAAACAAA TAGGATGAAACAAA TAGGATGAAACAAA TAGGATGAAACAAA TAGGATGAAACAAAAAAAAAA	TGGGGNANGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAAGACTCA ATCAGATAAT CCAATCCCAG GGAGACATAA TATTGAGAGG CAGAGTCTGA ATACAGATAGT TTTTGTTGGGA AAACAGTAGT TTTTGTTGGAA CTGATTAGA CTGATTAGAA CCTGATTAGA CTGATTAGAA CTTATTATAAA CCTGATTAGA TATTATAAAA CTTATTATAAA CTGATTAAAC TGAATAGA TGAATAGA TGAATAACT TACAGCATAA TGCAAATAGA TGATTAAAAT TGCAAATAGA TATTTAAAAT TGCAAATAGA TGATTAAAAT TATTTGGTGG TGACTTAAAT TATTTGGTGG TGACTTAAAT TATTTGGTGGG TGGGGTTAGAT TATTTGGTGGG TGGGGTTAGAT TATTTGGTGGG TGGGGTTAGAT TATTTGGTGG	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTOC TTATTGAGTT TGAGAGGTTT TTATCACTG GCTTTGCAAG GCATTGCAAG AGGCTAGAAC CAAATAGCTA AGGCTAGAC TATCATTTGAGG ATCCTGAATG TATCATATAC TATCAGCACGT AGGGCCATCG AGGATTCCT ACAATAGTCA AAGTAGGGC AGAATTCCT ACAATAGTCA AAGTAGGGC TACTGTTAAA CTATTTCTAA CAATAGTCA CAATAGTCA AAGTAGGGC CACTATTA CAACTATTA AAGCCATTTTA AAGCCATTTTA AAGCCATTTA CAACTATTA CGATCCATCG CGATCCATCG CGATCCATCG CGATCCATCG	ATAAGGATGA GCAATGTGAT CTTCTGCTATI ACAGTTTGG ACACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGA CCGTAATGAA CTGAACAGTA GAGTCTGGG GTTTTAGTATG GATCAGGAG CGGACTAGA CGGACTAGA CGGACTAGA CGGACTAGA CGGACTAGA TTGCATCAGT ACACTGAGGAG TTTCATTATA AACATGAAG CCTAATTACA CCTGATCTGGT CATAGCATT TATATCAGG CATAGCATT CATTTTGGT CATTTGGT CATTTGGT CATTTGGT CATTTTGGT CATTTGGT CATTGGT CATTTGGT CATTTGG	ACCTGGCTT TCTGCAGGCA TAAACTCTCA AATTTGAAA TGGCAGGTAC GATAAAATGT GATGAGTAGG ATTTGACAAA TGTTCAGGAG ATTTGACAAA TGACAGTACA ATTTGACAAA TAAAAAGAT TTAAACTGAC GAGTAAAAGAT TTAGACAAA TAAAAAGAT TAAAACAAA TAAAATGTGAC AAAATGTGAC AAAATGTGAC TAAATGTGAC TAAATGTGAC TAAATGTGAC TAAATGTGAC TAAAATGTGAC TAAAATGTGAC TAAAATGTGAC TAAAATGTGAC TAAAATGTGAC TAAAATGTGAC TAAAATGTGAC TAAAATGTGAC TAAAAAAAAATGGAC TAAAAAATTGACAAA TTTAAAGGTT TTTAGGAGTG ATTTTTTCC CCAAACAA TTTTTTTCC CCAAACAA TTTTTTTT	TEGAGCETEG TTGAGCEAG TTGGCTTCTA TTAMACAGT TTGGTTAGGG CCACTAGGTA CCTTCTCT TTTTGTATGG AMATGCAT TGGGAGGCG AMAGAGGGG TTGATGTATA AGCATGTAT AGCATGGATA AGCATGGATA ACCATGGATA ATTCAAGTGC TTAGAAGTAC TTAGAAGGCCA TTAGAAGTAC TTAGAAGTAC TCCAGGGCCA TCCACAGTCC TTCCACAGTCC GAAGGALATA	GAMTANTGA AGAGANTTAG ANAGGAGTTT TCTGTAMAGT TACTAGGGGT CTATAGGGGTT TTCCATATGG ATGGAGGTAT TCTGACAGA AGTTCAMATG TATGGACGATA TCAGCACAGA AGTTCAMATG TCAGCACAGA TCAGCACAGA TCAGCACAGA TCAGCACAGA TTAGAGCAC TCAGCACAGA TTAGGACGA AACAGCTTCT TCAGCACACAC AACAGCTTCT TCAGCACCACAC AACAGCTTCT TCAGCACCACAC AGGATTGTG TCAGCACCACAC AGGATATTTCC TCAGCACTAGA AGGATTCTGT TCAGCACCACC AGGAAACCTAG TGATATTTCC TCAGCACTAGA AGGAACCTAG AGGAAACCTAG TCAGTAATTGCC TCAGCACTAGA AGGAAACCTAG TCAGTAATTGCC TCAGCACTACC TCAGCACC TCACCACC T	1411 1511 1711 1911 2011 2211 2311 2411 2511 2611 2711 2911 3011 3111 3211 3311 3411 3511 3611 3711 4011 4011 4011
CIOTTGAGGA CGGTGAGTGA ATAGGGATAA ATAGAGAAAG CTGATATAAG ACTATGACAA ACTAGGAGAA CGAAGTATAAG ACTATGACAA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTATGC TGTACTTCC TGTACACTA TAGTACAGTA TAGTACAGTA ACAGGAGTTA ACAGGAGTA ACAGGAGAA ACAGAGAA ACAGAGAA ACAGAGAA ACAGAGAA ACAGAGAA	ACCIMANDA TITGTIAGA ACCAGACTCC GAACACATTA GGGATITTAT GGGATATTAA GGGATTCTGA ACTAAGGAG GCATTCCTGA AGTAGAGAAC CTAAGAGAAC TCATTATAA GCAGAACAT TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA AAAGCTGAG GAGCAAAGCG CATAAACTCA TAAAGAGCA TAAAGAGCA CATAAACTCA GATCTTGCCA AAGGTCACAG GCTCTCTTG CATCTTTTATC AAGTACAAAG GCTCTCTTTG CCCACAATT GAGCAGAGG GCCACCAATT GCAGCAGAGG GCCACCAATT GCAGCAGGG GCATCAAAGG GCCACCAATT GCAGCAGGG GCATCAAAGG GCCACCAATT GCAGCAGGG GCATCAAAGG GCCACCAATT	AGGATTATAC GGAATTATAC AATGGGAAA GCAAAGAAG TAACAATGGC TACCACTGAT GACTATAGAC TACCACTGAT GACTATAGAC TACACTGAT TACCACTGAT TACCACTGAT TACACTATAC TACACTATAC TACATATACA TACACTATAC TACATATACA TACACTATAC TACATATACA TACACTATAC GAGGACCAG AAATTATAC TACACTGTGA TACACTGTA TACACTGT	TAGGATHCEC TAAGGATAGA TICATGCTGC AGATGTCCTT TAAAGCAGCA ACACTGCTGA AGAGCTGGTC GGCTGATTTA ATTATCTGCG ACCATCTGGG GCTCAGTTAA ATTATCTGCTATTA TAGCTATTA TAGCTATTA GGGAGAACCC ATAGAGTGAT TTAGGTCATT GAGAGATATTA GGGAGATATTA GGCAGATACT AGCCCACAGAT ACCCCACAGAT TAGGTTGAT ACCCCACAGAT TAGGTTGAT ACCCCACAGAT TAGGTTGAT ACCCCACAGAT TAGGTTGAT ACCCCACAGAT TAGGTTGAT ACCCCACAGAT TAGGTTGAT ACCCCACAGAT TAGGTTGAT ACCCCACAGAT TAGGTTGAT GAGCCATCAG TCCCACATTAG GAGCCATCAA TCCCACATTAG GAGCCATCAA TCCCACATTAG GAGCCATCAA TCCCACATTAG GAGCCATCAA	TGGGGNANGG AGATANTATT TCACTTTAGG CATTCAGACA TGAGGCCAGG AGAGACTAA CAAAGACTCA ATCAGATAAT TGCTCTTAGG GAGACATAA TATGGAAGG CAGAGTCTAA AACAGTAGT TTTTTTGTGA CTGATTGTG AAACAGTAGT TATTATAAAA CCTGATTGTG AAACAGAAGA CTAATTAAAC TCAATTAACT TCAAATAACT TCACTTAAATAACT TCACTTAAATAACT TCACTTAAATAACT TCACTTATAATACT TATTATACT TCACTTATAATACT TCACTTATAATACT TCACTTATATACT TCACTTATAATACT TCACTTATATACT TCACTTATACT TCACTTATATACT TCACTTATACT TCACTTATAC	ATAGATAGA AAGGAGAAG GEAGCACTCT TTATTATTOC TTACTGAGTT TGAGAGGTTT TTATCATCTG GCTTTGCAAG GCTTTGCAAG GCTTTGCAAG AGGCTAGAAC CAATAGCACGT AGGCTAGAC TACAGCACGT AGGCCATGG AGGCTATGG AGAGTATACTC AGAGTAGAC AGATTACTCA AGATTACTCA AGATTACTCA AGATTACTCA ACACTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAACTACTAC CAACTACTAC CAACCACCTC TGCAGGGGGC CCCACCACTTC TGCAGGGGCACTT	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATAGGT GGTCTTAAAA AACGTGAGGT ACCTTAAGGAG CCGCTAATGGAG CTGTAAGAA CTGAACAGTA GATCTTAGGAG GATCATGGG GTTTTAGTATG GATCATGGG GTTTGGGCAT CCGGACTAGA AAGAAGGGAG AAAAAGGGAG AAAAAGGGAG AAAAAGGAG AATATTAAC CCTAATTTAA AACATGAAAG AATATCAAG CCTAATTTAC CATACGATT TATATCGGAG ATGGATCAGA CATTTTGGTG CCACAGCATT TCACAGCAGT TCACAGCAGCC CCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	ACCTGGCTT ACTGGCACA TAMACTCTCA AMTTTGANA TGCCAGGTAC GATANATGT GATGAGTAGG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG GCAGTACAAA GACTTGTGT TTACCTTGGCA GATANATGGAT CTGGAGAGTA CTGGAGAGT CTGGGAGAGT CTGGGAGAGT AGATCAACAA TAMATGTGAT GAGTAGGAGA ATTICAGCAT ANAMANTGG ACATTGAACAA TTTAAGGATA ACATTGTATT ACTTGAACAA TTTAAGGATA ACATTGTATT ACTTGAAGGAG ATTTTTTTCC GCAAACGAGT ATTTTTTCC GCAAACGAGT AGGCGGGAGAGAGAGACAAG AGGCGAGGCAG	TEGAGCETEG TTGAGAGCAG TTGGTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAGT GGAGGGGGG AAAAGAGGGG TTGAGTTAG AAACTGAAT AGCATGGAT ACCATGGAT ATCAAGTTAC TTAGAAATTAAGTTAC TTAGAAATTAAGTTAC TTAGAAATTA TAAGGAAGT ATTAAAGTTC TTAGAAATTA TAAGGAAGT ATCAAGTGC TTAGAAATTA TAAGGAAGT ATCATGAT AGCTATTAAG CCACAGATCT TCCAGGGCCA TTCACAGTCC GAAGGAAGT TCCACAGTCC GAAGGAAGT TCCACAGTCC GAAGGAAATTA TCCAGGGCCA TTCACAGTCC GAAGGAAATTA TCACTTGAG	GAMTANTGA AGAGATTAT TCTGTAMAG TACTAGGGT TTAGAGGATC CTTAMATGT TTCCATATAG ATGAGGTAT TCTGACATATG ATGAGGTAT TATGAGGTAT TATGAGGTAT TATGAGCACAT TATGAGCACAT TATGACACACAT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACAC TCTACTGAA ACCAGCATAA ACAGCTTCT TCAGCACACAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAATATTGC TCAACCACCAC TCAACACACC TCAACACC TCAACACACC TCAACACACC TCAACACACC TCAACACACC TCAACACACC TCAACACC TCAACC	1411 1511 1711 1811 1911 2011 2111 2211 2311 2411 2511 2611 2711 3011 3111 3111 3511 3511 3611 3711 4011 4011
CIOTTGAGGA CGGTGAGTGA ATAGGGATAA ATAGAGAAAG CTGATATAAG ACTATGACAA ACTAGGAGAA CGAAGTATAAG ACTATGACAA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTATGC TGTACTTCC TGTACACTA TAGTACAGTA TAGTACAGTA ACAGGAGTTA ACAGGAGTA ACAGGAGAA ACAGAGAA ACAGAGAA ACAGAGAA ACAGAGAA ACAGAGAA	ACCIMANDA TITGTIAGA ACCAGACTCC GAACACATTA GGGATITTAT GGGATATTAA GGGATTCTGA ACTAAGGAG GCATTCCTGA AGTAGAGAAC CTAAGAGAAC TCATTATAA GCAGAACAT TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA AAAGCTGAG GAGCAAAGCG CATAAACTCA TAAAGAGCA TAAAGAGCA CATAAACTCA GATCTTGCCA AAGGTCACAG GCTCTCTTG CATCTTTTATC AAGTACAAAG GCTCTCTTTG CCCACAATT GAGCAGAGG GCCACCAATT GCAGCAGAGG GCCACCAATT GCAGCAGGG GCATCAAAGG GCCACCAATT GCAGCAGGG GCATCAAAGG GCCACCAATT GCAGCAGGG GCATCAAAGG GCCACCAATT	AGGATTATAC GGAATTATAC AATGGGAAA GCAAAGAAG TAACAATGGC TACCACTGAT GACTATAGAC TACCACTGAT GACTATAGAC TACACTGAT TACCACTGAT TACCACTGAT TACACTATAC TACACTATAC TACATATACA TACACTATAC TACATATACA TACACTATAC TACATATACA TACACTATAC GAGGACCAG AAATTATAC TACACTGTGA TACACTGTA TACACTGT	TAGGATHCEC TAAGGATAGA TICATGCTGC AGATGTCCTT TAAAGCAGCA ACACTGCTGA AGAGCTGGTC GGCTGATTTA ATTATCTGCG ACCATCTGGG GCTCAGTTAA ATTATCTGCTATTA TAGCTATTA GGGAGAACCC ATAGGATGAT TTAGGTCATT GAGAGTATA GGGAGAACCC ATAGGATGAT GAGAGTATTA GGCAGATACT AGCCCACAGTA TAGGTCATA ACCCCACAGTA TAGGTTGAT ACCCCACAGTA TAGGTTGAT ACCCACAGTA TAGGTTGAT ACCCACAGTA TAGGTTGAT ACCCACAGTA TAGGTTGAT ACCCACAGTA TAGGTTGAT GAGGCATCAA TAGGTTGAT GAGCCATCAA TCCAATTAG GAGCATCAA	TGGGGNANGG AGATANTATT TCACTTTAGG CATTCAGACA TGAGGCCAGG AGAGACTAA CAAAGACTCA ATCAGATAAT TGCTCTTAGG GAGACATAA TATGGAAGG CAGAGTCTAA AACAGTAGT TTTTTTGTGA CTGATTGTG AAACAGTAGT TATTATAAAA CCTGATTGTG AAACAGAAGA CTAATTAAAC TCAATTAACT TCAAATAACT TCACTTAAATAACT TCACTTAAATAACT TCACTTAAATAACT TCACTTATAATACT TATTATACT TCACTTATAATACT TCACTTATAATACT TCACTTATATACT TCACTTATAATACT TCACTTATATACT TCACTTATACT TCACTTATATACT TCACTTATACT TCACTTATAC	ATAGATAGA AAGGAGAAG GEAGCACTCT TTATTATTOC TTACTGAGTT TGAGAGGTTT TTATCATCTG GCTTTGCAAG GCTTTGCAAG GCTTTGCAAG AGGCTAGAAC CAATAGCACGT AGGCTAGAC TACAGCACGT AGGCCATGG AGGCTATGG AGAGTATACTC AGAGTAGAC AGATTACTCA AGATTACTCA AGATTACTCA AGATTACTCA ACACTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAATTACTCA CAACTACTAC CAACTACTAC CAACCACCTC TGCAGGGGGC CCCACCACTTC TGCAGGGGCACTT	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATAGGT GGTCTTAAAA AACGTGAGGT ACCTTAAGGAG CCGCTAATGGAG CTGTAAGAA CTGAACAGTA GATCTTAGGAG GATCATGGG GTTTTAGTATG GATCATGGG GTTTGGGCAT CCGGACTAGA AAGAAGGGAG AAAAAGGGAG AAAAAGGGAG AAAAAGGAG AATATTAAC CCTAATTTAA AACATGAAAG AATATCAAG CCTAATTTAC CATACGATT TATATCGGAG ATGGATCAGA CATTTTGGTG CCACAGCATT TCACAGCAGT TCACAGCAGCC CCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCT TCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	ACCTGGCTT ACTGGCACA TAMACTCTCA AMTTTGANA TGCCAGGTAC GATANATGT GATGAGTAGG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG GCAGTACAAA GACTTGTGT TTACCTTGGCA GATANATGGAT CTGGAGAGTA CTGGAGAGT CTGGGAGAGT CTGGGAGAGT AGATCAACAA TAMATGTGAT GAGTAGGAGA ATTICAGCAT ANAMANTGG ACATTGAACAA TTTAAGGATA ACATTGTATT ACTTGAACAA TTTAAGGATA ACATTGTATT ACTTGAAGGAG ATTTTTTTCC GCAAACGAGT ATTTTTTCC GCAAACGAGT AGGCGGGAGAGAGAGACAAG AGGCGAGGCAG	TEGAGCETEG TTGAGAGCAG TTGGTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAGT GGAGGGGGG AAAAGAGGGG TTGAGTTAG AAACTGAAT AGCATGGAT ACCATGGAT ATCAAGTTAC TTAGAAATTAAGTTAC TTAGAAATTAAGTTAC TTAGAAATTA TAAGGAAGT ATTAAAGTTC TTAGAAATTA TAAGGAAGT ATCAAGTGC TTAGAAATTA TAAGGAAGT ATCATGAT AGCTATTAAG CCACAGATCT TCCAGGGCCA TTCACAGTCC GAAGGAAGT TCCACAGTCC GAAGGAAGT TCCACAGTCC GAAGGAAATTA TCCAGGGCCA TTCACAGTCC GAAGGAAATTA TCACTTGAG	GAMTANTGA AGAGATTAT TCTGTAMAG TACTAGGGT TTAGAGGATC CTTAMATGT TTCCATATAG ATGAGGTAT TCTGACATATG ATGAGGTAT TATGAGGTAT TATGAGGTAT TATGAGCACAT TATGAGCACAT TATGACACACAT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACAC TCTACTGAA ACCAGCATAA ACAGCTTCT TCAGCACACAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAACCACCAC TCAATATTGC TCAACCACCAC TCAACACACC TCAACACC TCAACACACC TCAACACACC TCAACACACC TCAACACACC TCAACACACC TCAACACC TCAACC	1411 1511 1711 1911 2011 2211 2311 2411 2511 2611 2711 2911 3011 3111 3211 3311 3411 3511 3611 3711 4011 4011 4011
CCOTTGAGGA CGGTGAGTGA ATAGGGATAA ATAGAGAAAG CGAAGTATAAG ACTATGACAAA CGAAGTATAAG ACTATGAGAAA CGAAGTATAA CGAAGATAA TTAGGGAAA TTATAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA TAGTACTTC CTCCAACACA TATCTGACTA TGAAAAAGGA ACAGGAGTTA ACGGGGTCTT TGCTGAGCA GCGAAGGAA GCCAATGAAG CCCATGAAGAG CCCATGAAGAG CCCATGAAGAG CCCATGAAGAG	ACGGGCCAGG TITGCITGAGA TITGCITGAGA ACCAGACTCC GAACACATTA GGGATTITAT GTGAGACAGG GCATTAAGGAG GCAGTCCTGA AGTAGAGAAC TCATTITATAA GCAGAACCAT TAGGTGAGTA GAGCAGTAGAGGA CATAAGGAG GAGCAACAGAGCAC TAATITATAA GCAGGAGCAT TIAAAGACTAG GAGCAAACCAG CATAAACTCAG CATCTCCTGC AAGGTCACAG CATCTCCTTA AAGGTCACAG GCTCCCTTG AAGGTCACAG CTCCTTTTATC AAGGTCACAG GCTCCCTTG AAGGTCACAG CTCCTTTTATC AAGGTCACAG CTCCTTTTTATC AAGGTCACAG CTCCTTTTTATC AAGGTCACAG CTCTTTTTATC CTCTTTTTATC CTCTTTTTATC CTCTTTTTATC CTCTTTTTATC CTCTTTTTATC CTCTTTTTTTATC CTCTTTTTTTT	AGGITAGTA TATGCATT TATGCATT TATGCATT TATGCATC ANTGGGANA GCANAGAGG TANCACTGAT GACTATTGCA TATCACTTA ANATAGACT TACACTATA ANATAGATC TACATATACA TITTGGACTT ACACTATACA TACATATACA TACATATACA TACATATACA TACATATACA TACATATACA TACATATACA TACAGTACTA ANAGAGATT TACACTGGA GTCTAGAGGA TCAGTTCATA TACACTGTGA ACATGCGCTGCA AGGTTGAAATG	TAGGATHCECE TAAGGATAGA TTCATGCTGC AGATGTCTTT TAAGGATAGA ATCATCGTG AGAGCTGGTTA ATTATCTGCG ACCATCTGGG GCTCATTTA ATTATCTGCG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ATTATCTGCT ATTATCTGCT ATTATCTCTT GCGAGAACCC ATAGGTGAAA TAGGAGTAAT TTAGGTCATT GAGAAATTTA GGCAGATACT ACCCACAGAT TAGGTTATA ACTGAGGAAC TAGGTGAT TAGGTTATA ACTGAGGAAC CACCACTCAA TAGGTTATA TAGGTTATT TAGGTTATT TAGGTTATT ACCATTAACAA TTAGGTTATT TAGGTTATT TAGGTTATA TAGGTTATA TAGGTTATA TAGGTTATA TAGGTTATA TAGGTTATA TCCATTAGGAAC GACCCATCAA TCCAATTTAG TGGTGGCTCA GTGAAACCC GTGAAACCC	TGGGANAGG AGTANTATT TCCTTTAGG CATCTAGCA TGAGGCAGG TGAGGCAGG ACAGAATAAT TGCTCTAGACA ATCAGATAAT CANAGATCA ATCAGATAAT TTTGTAGGA CAGAGACTAA AACAGTAGT TTTTGTTGAG TGTTGTTGAA TATTATAAA TCTGATTGTG AAACAGAAGA TATTATAAAA TCTGATTGTG AAACAGAAGA TATTATAAAA TCTGATTGTG AAACAGAAGA TGATTCAAC TACAGATAA TGCAAATAGT TACAGCATAA TGCAAATAGT TACAGCATAA TATTGTGTGG GGGCTAGGAT CACCTTAATT CACCTTAATT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTCC TTACTGAGTT TTATCTG GCTTTGCAAG GCAGCACA AGGCTAGAAC GAATCATTA AGGCTAGAAC CTATTTGTAGG CTATTTGTAGG TATCATATAC TACAGCACGT AGGGCCAGGG AGGCTATGG AGACTATCCT ACATTAGTCA ACATTAGTCA ACATTAGTCA CAACTATT	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATAGGT GGTCTTAAAA AACGTGAGGT ACCTTAAGGAG CCGCTAATGAA CTGAACAGTA CATGATGAG CTGAACAGTA CATGATGAG GTTTAGAATGAA CTGAACAGTA CAGACAGTA CAGACAGTA AACAGGAG AAAAAGGGAG ATAAAAGGAG ATAAAAGGAG CCTAATATAA ACATGAAAA CCTGATCTGAT	ACCTGGCTT TCTGCAGGCA TAAACTCTCA AATTTGAAA TGCCAGGTAC GATAAAATGT GATGAGTAGG ATTTGAGAGA ATTTGAGAGA ATTGGACAAT TTACACTGA GCAGTACAAA GACTTTGGTGT TTACCTTGGA GATAAAGAT TTAGGAGAGGAG	TEGAGCCTEG TTGAGCAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAGT GCAGGGGGG AAAAGAGGGG AAAAGAGGGG TTGAGTTAGA AGCATTAGAAT AGCATGGATA AGCATGGATA ATTCAAGTGC TTAGAAATTAA AGCATGGATTA ATTCAAGTGC TTAGAAATTA AGCATGGATTA ATTCAAGTGC TTAGAAATTA AGCATCAAGTGC GTATTTTCAC ATGGTTCGAT ATCAAGGTC GTTTTTTCAC ATGGTTCGAT TCCAGGGCCCA TTCCAGGGCCCA TTCCAAGGTCC GAAGGAAATT TCCAAGGCCCA TTCCAAGGCCCA TTCCAAGTCC GAAGGAAATT TCCAAGGTCC GAAGGAAATT TCCAAGGTCC GAAGGAAATT TCCAAGGTCC GAAGGAAATT TTCACAGTCC GAAGGAAATT TCACAGTCC GAAGGAAATT TCACAGTCC GAAGGAAATT TCACAGTCC GAAGGAAATT TCACAGTCC GAAGGAAATT CCCTCCTCCTC CCCCTCCTCCTCC CCCCCTCCTCC	GAMTANTGA AGAGAGTTT TCTGTAMAG TACTAGGGT TTCATATGG TTCATATGG ATGGGGTATTCCTACATGG ATGGGGTAT AGTCATATGG ATGGAGGATC TATGGACACA AGTTCAAATG TCAGCACAGA TCAGCACAGA TCAGCACAGA AGAGTATAAAGG TGGATGGATAAAAGG TCAACCACACACACACACACACACACACACACACACACA	1411 1511 1711 1811 2011 2211 2211 2211 2511 2611 2711 2911 3011 3111 3211 3511 3511 3511 3511 4011 4011 4114 4314 4314
CCOTTGAGGA CGGTGAGGA CAGTTGAGTAA ATAGAGAAAG CTGATATAAG ACTATGAGAAAG CTAATATAAG ACTATGAGAGA ACTATGAGAGA ACTATGAGAGA ATATTAAGGA AGAATATATG TGTATCTTCC CTCCAACACA TATTTAGGA TATATGAGTA TAGTACATTG TAGAAAAGGA ACAGTAGATA ACAGGAGTTA ACAGGAGTTA ACAGGAGTTA ACAGGAGTTA ACGGAGTAT ACGGTGATCA CTGTACAGTC TGTACAGTC TGTACAGTC TGTACAGTC TGTACAGTC TGTACAGTC TGTACAGTC TGTACAGTC TGTACAGTC TGTACAGTC TGTACAGC ACTTCAGGC ACGGAGGAGA ACCCAATGAAG CCCAATGAAG CCCAATGAGAG ACCCAGC ACTTCGGAGG ACTTCGGAGGAGGAGA ACCCAATGAAG ACCCAATGAAG ACCCAATGAAG ACTTCGGAGGG ACTTCGGAGGG ACTTCGGAGGGAGGAGGAGAGAA ACCCAATGAAG ACCCAATGAAG ACTTCGGAGGGAGGAGGAGAA ACCCAATGAAG ACCTCATGGAGG ACTTCGGAGGGAGGAGGAA ACTTCGGAGGAGGAGAAA ACCCAATGAAGAGG ACTTCGGAGGAGGAGAAA ACCCAATGAAGAGGAGAAAA ACTTCGGAGGAGAAAAA ACTTCGGAGGAGAAAAAAAAAA	ACGGGCCAGG TITGCTGAGA TITGCTGAGA ACCAGACTCC GAACACATTA GGGATTTTAT GTGAGACAGG GCATTATAGGAG GCAGTCCTGA AGTAGAGGAC CTAAGAGAAC TCATTTATAA GCAGAACCATTATAAA ACAGAGCAGAG	AGGITMETT GGAATTITTC GGTTTGCATT PATGCATCA ATTGGGGAA GCAAGGAGG TAAACTAGGC TACCACTGAT GACTATGCA GTATGACTTA AAATAGATG AAATAGACT AAATAAGATG TACATATACA CTGGTTGTAT AAATAAGACC CTGGAGCCT AAAGAGACT AAAGAGACT AAAGAGACT AAAGAGACT AAAGAGACT AAAGAGACT AAAGAGACT ACCCTGGA TCGATGCAT TCCATGGAT TCCATGGAT TCCATGGAT TCCATGGAT TCCATGGAT TCCATGGAT TCCATGGAT AGAGGCTGT AGAGGCTGT AGAGGCTGT AGAGGCTGT AGAGGCTGCA AGGCTGCAA	TACGATICCCC TAAGGATAGA TTCATGCTGC AGATGTCTTT TAAAGCAGCA CAATTGATAG AGACCTGGTC AGACCTGGTC ACCATCTGGG ACCATCATTAAATAT GGGAGAACCC ATAGAGTGAT TAAGATTTA GATAAATAT GAGAAATTTA ACTAGAGTAA TTAGGTTATA ACTAGAGTATA ACTAGAGTAT TAGGTAAACA TTAGATTAA TTAGGTTATA ACTAGAGTAT TAGGTTATA ACTAGAGTAT TAGGTTATA ACTAGAGTAC TCCAATTTAG TGCAATTTAG TGGAACCCC GAACCTGGGA	TGGGGNAGG AGATAATACT CAGTATTAAT TCCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAATAA ACAGAGATAA TATGGAAGGC GGAGACATAA TATGGAAGGC CAGAGTCTGA CAGAGTCTGA AAACAGTAGT TATTGTTGAC TGGACATTA TATTATAAAA CCTGATTGTG AAACAGAGAG TCTATTATAC TGATTCAAC TCAATTAAC TCAATTAAC TCAATTAAC TCAATTAAC TCAATTAAC TCAATTAAC TCAATTAAC TCAACTAGAG TAATTGGTGG GGGGTAGGAT CACCTATAAT CACCTATAAT CACCTATAAT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTOG TTACTGAGTT TTATCTGAGTT TTATCATCTG GCTTTGCAAG GCATTCATAA AGGCTAGAAC CAATAGCACA AGGCTAGAAC CATTTGCAAG CATTTGCAAG CATTTGCAAG AGGCTCACGT AGGCTCACGT AGGCTCACGT AGACTATCCA AGACTATCCA AGACTACCACCT TACAGCACCATCT GCAGGACCATCT TGCAGGGTCG CATCCATCTG TGCAGGGTCG CATCCATCTG TGCAGGGTCG CCACCACTCT TGCAGGGTCG CCACCACTCT TGCAGGGTCG CCCAGCACTT AAAATACAAA GCAGTCAGCC	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTTAAGGAG CCGTAATGAA CTGAACAGTA GAGTCTTAGGA GTTTTAGGAG GTTTAGGAGT TAGAGAGTAAAGTAA AGTTAGGGAG CGGACTAGAA AGGGAACAG AAAAAGGGAG CCGGACTAGAA AACAGAAAAGGGAG CCTAATTTACATC AACAGAAGCATTTAAATTCAAG CCTAATTTCAAT AACATGAAAG AATATTCAGA AATATTCAGA AATATTCAGA AATATTCAGA CATAGGATGTT TATATCGGAG CATAGGATGTT TATATCGGAG GCACAGCATT TCAGAGGCCC ÄATTAGCTGG GACATCGCAC	ACCTGGCTT ACCTGGCA TANACTCTCA ANTITIGANA TGCCAGGTAC GATANAGTGT GATGAGTAGG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG GCAGTACAAA GACTTIGGT TTACCTIGCA TTACGTIGGA TTACGGAGGT CTGGGAAGGT CTGGGAAGGT CTGGGAAGGT AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA TTAAAGTCT AAAAAATGGA ATTITACGAGT TTAAGAGTT ACTTGACAAA TTAAGAGTT TTAAGAGTT ACTTGACAA TTTAAGAGTT ACTTGACCAG GCAACCAGGCAG GCCAGGCAG GCCAGGCAG CCATGGTGGC CACTGCACCT CAAACAGCT CCACTGCACCT CAAACAGCAG GCCAGGCAGC CCATGCACCT CAAACAGCAGCAG CCATGCACCT CAAACAGCAGCAG CCATGCACCC CACTGCACCT CAAACAGCAGCAC CCATGCACCC CACTGCACCC CACTGCACCC CCATGCACCC CCATGCACCC CCATGCACCC CATGCACCC CCATGCACCC CATGCACCC CATGCACCC CCATGCACCC CATGCACCC CATGCACC CATGCACCC CATGCACC C	TGGAGCCTGG TTGGAGCCAG TTGGCTTCTA TTAAAACAGT TTGTGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAGT CGCAGGGCGC AAAAGAGGGG TTGAGTTAC AAACTGAAT AGCATTAC AATCGAAGT AATCGAAGT ATCAAAGTTAC AATCGAAGT ATTCAAGTGG TTAAGATTAC ATAGAAGTA TTAAGAATTAA GTATTGAAA TTAAGATTC TTAAGATTCA ATCGTTCGAT TCCAGGGCCA TTCACGTCGT TCCAGGGCCA TTCACGTCGAT TCCACGGCCA TTCACAGTCG TCCAGGGCCA TTCACAGTCG TCCACGGCCA TCCACGGCCA GCGAGGAAATTAA GCATTCACTTCGAT TCCACAGGCCA GCGAGGAAATTAAG CCACAGGCCA GCGAGGAAATTAAG CCACAGGCCA TCCACGGCCA TCCACGGCCA GCGCCTGGC CACGCCCTGGC CACGCCCTGGC CACGCCCTCGC CCACGCCCTCGC CCACGCCCTCT TCCACGGCCA CCCCCCCCCC	GAMTANTGA AGAGAGTAT TCTGTAMAG TACTAGGGGT TACTAGGGGT TTCCATATGG ATGGAGGTAT CCTACAAG ATGGAGGTAT CCTACAAG ATGGAGGTAT TATGGACCAC CCTATGAA TCAGCACAG TATGGACCAC TCAGGACCAC TCAGGACCAC TCAGGACCAC AGGCATAAA ATATATAAGG TGGATGGTAT TTTAGACCAC AACCACCAT AACCACCAT AACCACCAC TCTACTGAT TCAACCAC TCTACTGAT ACACCTTCT TCAACCAC TCTACTGAT TCAACCACC TCTACTGAT TCAACCACC TCTACTGAT TCAACCACC TCTACTGAT TCAACCACC TCAACTGAT TCAACCACC TCAACTGAC TCACCACC T	1411 1511 1711 1811 2011 22111 22111 22111 2311 2411 2511 2611 2711 3011 3011 3111 3511 3511 3511 4011 4111 4311 4411
CCOTTGAGGA CGGTGAGGA CAGTTGAGTA ATAGAGAAAC CGAAGTATAAG ACTATGACAA CGAAGTATAAG ATATGACAA CGAAGTATAAG ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTAAGGA ATATTATCTTCC GTCCAACCA TATTTAGCTA TAGTACAGTA TAGTACAGTA ACAGGAGTTA TGAATAAGA ACAGGAGTTA TGACTAGTA ACAGGAGTTA TGACTAGTA ACAGGAGTTA TGACTAGTA ACAGGAGTTA TGACTAGTA ACAGGAGTTA ACGGGGGTTT TGACTAGTA ACAGGAGTTA ACGGGGGTTT TGACTAGTA ACGGGGGTTT TGCTCACGT TGACAGTA ACTGACAGGA ACTGACAGGA ACTGACAGC ACTTGACAGA CCCAACCAGC ACTTGGGAG ACTTGGGAG ACTTGGGAG ACTTGGGAGG ACTTGGAGGAAA ACATTGGAGAGAA ACATTGACACAC ACTTGGAGGA ACTTGGGAGG ACTTGGGAGG ACTTGGGAGG ACTTGGGAGG ACTTGGGAGG ACTTGGGAGG ACTTGGGAGG ACTTGGAGGAAACA ACTTGGGAGG ACTTGGGAGG ACTTGGGAGG ACTTGGGAGGAAACACACAC ACTTGGAGGAGAACACACACC ACTTGGAGGAGAACACACACC ACTTGGAGGAGAACACACC ACTTGGAGGAGAACACACC ACTTGCACACCACC ACTTGCACACC ACTTCCACACC AC	ACGGGCAGG TTGCTGAGA TTGCTAAAGTG ACCAGACTCC GAACACATTA GGGATTTTAT GTGAGACAGG ACATANGGAG GCAGTTGAAAGTGA AGTAGAGAC TCATTTATA GCAGAACGCAT TAGTGGAGAC CTATTATAA GCAGAACCAT TAGTGGAGAC CTGGCAGTA AAAAGCTGAG GAGCAATGCAG GATCATGCCA AAAACCACA CTCTTTTATC AAGGCAAAGC GCTCTCCTTG CTCTTTTATC AAGGCAAAGC CTCTTTTATC AAGGCAAAGC CTCTTTTATC AAGGCACACATT GCACACACATT GCACACACACAC CTGGCCACA CTGGCCACA CTGGCGACA CTGGGCACA CTGGCCACA CTGGGCACA CTGGGCACA CTGGGCACA CTGGGCACA CTGGCCACA CTGAGGCACAC CTGGCCACA CTGAGGCACAC CTGAGCACAC CTGAGGCACAC CTGAGGCACAC CTGAGGCACAC CTGAGCACAC CTGAGGCACAC CTGAGCACAC CTGAGGCACAC CTGAGGCACAC CTGAGACACAC CTGAGGCACAC CTGAGGCACAC CTACACAC CTGAGGCACAC CTGAGGCACAC CTGAGACACAC CTGAGGCACAC CTGAGACACAC CTGAGGCACAC CTGAGACACAC CTGAGGCACAC CTGAGACACAC CTGAGACACAC CTGAGACACAC CTGAGACACAC CTGAGACACAC CTGAGACACAC CTGAGACACAC CTGAGACAC CTGAGACAC CTGAGACAC CTGAGACAC CTGAGCACAC CTGAGCAC CTGAGCACAC CTGAGCACAC CTGAGCACAC CTCACAC	AGAITMETT GGAATTTTC TOTTGCATT PATGCATCA CTCTTTGATC AATGGGAAA GCAAGAAGG TACCACTGAT GACTATGCA TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCATTGAT TACCATTGAT TACCATTGAT TACCATTGAT ACTACTGAGACTC GCGGAACCAGA AAATTATTCT ACACCGTGGAA TCCATCGGAG TCCATCGAGA TCCATCGAGA TCCATCGAGA TCCATCGAGA TCCATCGAGA TCCATCGAGA TCCATCGAGA TCCATCGAGA TCCATCGAT TCCATCCAT TCCATCCAT TCCATCGAT TCCATCCAT TCCATCCAT TCCATCCAT TCCATCCA	TAGGATAGA TTAAGGATAGA AGATGTCTT TAAAGCAGCA AGATGTCTT TAAAGCAGCA AGATGTCTGTG AGAGCTGGTC AGAGCTGGTC ACCATCTGGG GTCTGACTGC ACCATCTGACA ATTACTATTA ATTACTATTA ATTACTATTA GGTAGATACA TTAAGATTTAATAT TAAGGTGATT AGAGATTTAACAT TTAAGTATT AGAGATTAA CTTAAATAT AGTAGAT TAAGGTGAT TAAGGTGAT TAAGGTGAT TAAGGTGAT TAAGGTGAT TAAGGTGAT TAAGATTAA ACTAGGTAA ACTAGGTAACAA ACTAGGGAAC GACCATTAACAA TTAAGGTATT AGCCACTAACAA ACTAGGGAAC GACCATCTAACAA TTAAGGTATT AGTAGGTAAC ACTAGAGTAC TGGGGAAC GACCATCTAACAA TGGATACATTAAG TGGTGGCTCA GACCATCTAACAA TGGATAGATTAA ACTAGGGAAC GACCATCTAA GAGGGAAC GACCATCTAA GAGGGAAC GACCATCGGGAA AAAAAGTCTA	TGGGAMAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAATAC ATCAGATAAT CGATGCCCAG GGAGACATAA ATCAGATAAT TTTTGTGGAG CAGAGTCTGA ATACGAGATAA AAACAGTAGT TTTTGTTGAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTATTTTAAAA CCTGATTTGTG CACTGATTAAC TCAATTAAC GGGTTAACT GGGTTAACT GGGTTAACT GGGTTAACT GGGTTAACT GGGTTAACT GGGTGGAGGT GGGTGGAGGTT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTATTGAGTT TGAGAGGTTT TTTTCATCTG GCTTTGCAAG GCATTGCAAG GCATTGCAAG AGGCTAGAC AGGCTAGAC ATCCTGAATG CTATTTTTTTTTT	ATAAGGATGA GCAATGTGAT ACAGTTTCG ACACTACGG CAATATAGGG CAATATAGGT ACCGTAAGGAG ACCTAAGGAG CCGTAATGAA CTGAACAGTT GATTAGGAG TITTAGTATG GATTAGGAG TITTAGTATG ACTTAGGAG TITTAGTATG ACTTAGGAG TITTCATTAGA TGGTTCACTC AGAGGAACAG AATAAGGGAG CCTAATTTAC CTCATTTGAT TATATCAAG CATAGCATTT TATATCAAG CATAGCATTT TATATCAAG CATAGCATTT TATATCAAG CATAGCATTT TATATCAAG CATAGCATTT TATATCAAG CATAGCATT TATATCAAG CATAGCATC AATTAGCTGG GCACAGGACC AATTAGCTGG GGAGATCGCAC AGTGGATCCCA	ACCTGGCTT TCTGCAGCA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA TGGCAGGTAC GATANATGT ATGTCAGAGA ATTCGACAAT TAGATAGTGG GCAGTACAAA GACTTGTGT TTACCTTGCA GLTANATGGT TTACCTTGCA GLTANATGGACAT TTAGAGAGAT TTGGAGAGAGT TTGGAGAGAGT TTGGAGAGAGA	TEGAGCCTEG TTGAGCCAG TTGGCTTCTA TTAMACAGT TTGTCTAGGG GCACTAGGTA TTTTCTATGG TTGTTAGGG GCACTAGGTA AMATGCAT TTTTCTATGG AMATGCAT TTGTTAGGG TAGAGGGGG AAAGAGGGG TAGAGGTGA AACCCCTTAG ACCCCTTAG ACCCTTAGAAT ACCATGGATA ACCATGGATA TTAGAGTAC TTAGAATTAA TAAGGAAGA ATAAGGATCT TTAGAATTAAGGATCA TTAGAATTAAGGATCA TTAGAATTAAGGATCA TTAGAATCA TTAGAATTAAGGATCA TTAGAATTAAGGATCA TTAGAATTAAGG CCACAGATCT TCCAGGGCCA TTCACAGTCC GAAGGAAATT ATCACTTGAG CCACCGGCCCTGT CACCCTGGGCCA CCACCCTGGGCCAGAGCAAGACT CCACCGGCCCTGT CACCCGGGCCAGACCT CCACCGGCCCAGACCT CCACCGGCCCAGACCT CCACCGGCCCAGACCT CCACCGGCCCAGACCAGA	GAMTANTGA AGATANTAGA AAAGGAGTTT TCTGTANAAT TACTAGGGGT CTATAAGGGTT TTCCATATAG ATTCCATATAG ATTCCATATAG TCTACAAGA AGTTCANATG TATGGAGGTAT TCAGGAGGACAG TTAGAGGACAG TTAGAGACAG TCAGGACAGAT TCAGGAGAGAT TCAGGAGAGGAT ATTAGAGAGGAT ATTAGAGAGGAT AAAGAGCTAC AAAACAGAT AAAACAGAT AAAACAGAT AAGAGCTTCT TCAGGAGTTCT AATCCCAGGT AATCCCAGGT AAACCAGAGTT AATCCCAGGCT AAACCAGAGTT AATCCCAGGCT AATCCCAGGCTACT GAGAGAGCAA GAGAGCCAAG	1411 1511 1711 1811 2011 2211 2211 2311 2411 2511 2611 3011 3011 3111 3311 3411 3511 3611 3712 3811 4011 4111 4211 4311 4411
CCOTTGAGGA CGGTGAGTGA ATAGGGATAA ATAGGGATAA ATAGGGAAAA CGAAGTATAAG ACTATGACAA ACTATGACAA ACTATGACAA ACTATGACAA ACTATGACAA ACTATGACTA TGATCTTCC CTCCAACACA TATCTGACTA TGATATGACTA TGAAAAGGA ACAGGAGTTA ACAGGAGTCA ACTTGACGA ACTTGACGA ACTTGACGA ACTTGAGGA ACTTGAGAA ACAGGAGGA ACTTGAGAA ACAGGAGAA ACAGGAA ACAGGAAA ACAGAAA ACAGGAAA ACAGAAA ACAGGAAA ACAGAAA ACAGGAAA ACAGAAAA ACAGGAAA ACAGAAA ACAGGAAA ACAGAAA ACAGGAAA ACAGGAAA ACAGAAA ACAGGAAAA ACAGGAAAA ACAGAAAA ACAGAAAA ACAGAAAA ACAGAAAA ACAGAAAA ACAGAAAA ACAGAAAAA ACAGAAAA ACAGAAAAA ACAGAAAA	ACGOGCEAGG TITGCITGAGA TITGCITGAGA ACCAGACTCC GAACACATTA GCGAACACATTA GCGAATTTAT GCGAGACACG CTAAGGAG GCACTCCTGA AGTAGAGACC TCATTATAAA GCAGAACCAT TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA AAAGCTGAG GACCAACTACACAC TCATTITATAA GAGCACACAC TAAAGACCAC TAAAGCTCACA AAGGTCACAG GCACCACACT CACTAGAACACAC CTGAGGCACAC CACACACACACACACACACACACACACACA	AGUITMETT GGAATTITTC TGTTTGCATT PATGCATTCAT ATTGGGAAA CTAGGG TAACACTGAT GACTATTGCA TTAGACTTA AAATAGGT TACACTTTA AAATAGATT AAATAGATT TACATTACA TTTGGACTT AATAGACTC CTGGAGCCT CTGGAGCCT AAAGAGATTCAC AAATAGATTCA AAATAGATTCAC AAAGAGATTCAC TTCATAGTTTATT TACACTGGA TCAGTTCAT TTCATAGT TTCATAGT TTCATAGT TTCATAGT TTCATAGT TTCATAGT TCGAGCTGCA AGGTTGAAAA AGGTTGAAAA AGGTTGAAAAA AGGTTGAAAAA	TAGGATTICEC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAGGATAGA ATCATCGTG AGAGCTGGTC GCTGATTTA ATTATCTGCG ACCATCTGGG GCTCACTTA ATTATCTGCG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ATTGTTCTGT TTAGGTCATA TGGGAAACCC ATAGGTGAAA GGAAAATTTA GGGAAATTTA GGCAGATACT TAGGTTATA TCCATTAGA TCCATTAGA GTGAACCC GAACCTGGGA AAAAGTCCC GAACCTGGGA AAAAGTCCC GAACCTGGGA AAAAGTCCC	TGGGANAGG AGATANTATT TCACTTTAGG CATCTAGACA TGAGGCAGG AGAGACATAA CAAAGACTCA ATCAGATAAT TAGGCCAGG GGAGACATAA TATGGAAGG CAGGCCAGA TTTGTGGGAA AACAGTAGT TTTTTTGTGAC TGTGCCTAC TGGACATTAA AACAGTAGT TTTTTTTGTGA AACAGTAGT TATTATAAAA CTTATTATAAA TCTGATTGTG AAACAGAAGAG TCAATTAAC TCAATTAACT TACAGCATAA TACAGCATAAT TACAGCATAAT TACAGCATAAT TACAGCATAAT TACAGCATAAT CACCTATAAT CACCTATAAT GGCTTACTA GGCTTACTA	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT TTTTCATCTG GCTTTGCAAG GCATCCATCA AGGCTAGAAC CGAATCGATA AGGCTAGAAC CTTTTTGTAGG TATCAGCACGT AGGCTAGGAC AGGCTAGGAC AGGCTAGGG AGGCTAGGG AGGCTATGG AGGCTAGGG AGGCTATGG AGGCTATGT AGGCCATCG AGATCATA CTATTTCTGTAG CTATTTTTTA AGGCCATCTG TGCAGGTCGC CCCACCTG AAATACTAC CCCACCACTT AAATACAAAA CTATTTCTGT TGCAGGGTCG CCCACCACTT AAATACAAAA CTATTTCTGT CCCACCACTT CGCAGGTCGC CCCACCACTT AAATACAAAA CTAGGAAGA AAAGAAAAAA TAACGAAAAA	ATAAGGATGA GCAATGTATE CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTTAGGAG CCGCTAATGAA CTGAACAGTA GATCTTGGG TTTTAGTATG GATCATGGG GTTTGGGCAT GAACAGTA AAGTAGGAG ATGAACAGTA AAGTAGGAG AAAAGGGAG TTCATTATA AACATGAAG AATATCAAG CCTAATTAC CTGATCTGAT	ACCTGGCTT TCTGCAGGCA TAAACTCTCA AATTTGAAA TGCCAGGTAC GATAAAATGT GATGAGTAGG ATTTCAGGAG ATTTCAGGAG ATTTCAGGAG ATTTCAGAGAA GACTTTGTGT TTACCTTGGCA GATAAATGT TTACCTTGGCA GATAATGGAT TACATGGAGTAG GATTAATGAA TCTGGAGAGT CTGGGAGAGT CTGGGAGAGT CTGGGAGAGT AGATCAACAA TAAATGTGAT ACTTGAGAA TCTAAAGTCT TAAAAAATGTGAT ACTTGAGAGAA TCTAAAGAT TTTAGGAGAA TCTAAAGATT TTTAGGAGAA TCTAAAGATT TTTAGGAGAA TTTAAAGTCT TTTAGGAGAA TTTTTTTCC GCAAACAGGT ACGCAGGCAG GCATGGGCAG GCATGGGCAG CACTGCACTC AAAGGAACTAC CACTGCACTC AAAGGAACTAC CACTGCACTC AAAGGAACTAC CACTGCACTC CACTGCACT CACTGCACTC CACTC CACTGCACTC CACTGCACTC CACTGCACTC CACTGCACTC CACTGCACTC CACTGCA	TEGAGCCTEG TTGAGCCAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAGT AGGAGGGGGG AAAAGAGGGG TTGAGTTAG AAACTGCAT AGCATTAG CCACAGAGTTAC TTAGAAATTAAG TTAGAAATTA AGCATGGAT ATTCAAGTGC TTAGAAATTA TTAGAAGTTAC TTAGAAATTA TTAGAAGTT AGCTTCTTCACAGTCC GAAGGAAGT ACCACAGATCT TCCAGGGCCA TTCACAGTCC GAAGGAAGT ACCACAGTCC GAAGGAAGT ACCACAGTCC GAAGGAAGT ACCACAGTCC GAAGGAAGT ACCACAGTCC GAAGGAAGT CCACAGGCCA TTCACAGTCC GAAGGAAGT ACCACAGTCC GAAGGAAATTA CCACGGCCCA TCCACAGTCC GAAGGAAATTA CCACGGCCCA CCACGATCC CAGCCCTGGC CAGCCCTGGC CAGCCCTGGGC CAGCCCTGGGC CAGCCCTGGGC CAGCCCTGGGC CAGCCCTGGGC CAGCCCTGGC CAGCCCTGGCC CAGCCCTGCC CAGCCCTGGCC CAGCCCTGGCC CAGCCCTGCC CAGCCCTGCC CAGCCCTGGCC CAGCCCTGGCC CAGCCCTGCC CAGCCCTGCC CAGCCCTGCC CAGCCCTGCC CAGCCCTGCC CAGCCCTGCC CAGCCCTGCC CAGCCCTCC CAGCCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCCC CAGCCCCCCC CAGCCCCCC CAGCCCCCCC CAGCCCCCCC CACCCCCCC CACCCCCCCC	GAMTANTGA AGAGATTAT TCTGTAMAG TACTAGGGT TTAMAGGGT TTCATATAG ATGGAGGTAT TCTGATATAG ATGGAGGTAT TATTGAGCACA AGTTCAARTG TATTGACACACACACACACACACACACACACACACACACA	1411 1511 1711 1911 2011 22111 22111 22111 2311 2411 2511 2611 2711 2611 2711 3111 3211 3311 3311 3411 3511 3511 4011 4011 4411 4511 4611
CCOTTGAGGA CGGTGAGGA CGGTGAGGA ATAGAGAAAG ACTATGAGAAAG ACTATGAGAAAG ACTATGAGAGA ACTATGAGAGA ACTATGAGAGA ACTATGAGAGA ACTATGAGGAGA ATATTAAGGA ACTATGACTA TGATCTTCC CTCCAACCA TATCTGACTA TAGGAGAGTTA ACAGGAGTTA ACAGGAGTAAA CCACTACAGC CCAACGAGG CCAACGAGG CCAACCAGG CCACTCCAGC CCACTCCACT	ACCIMANDA AGGGGCAGG TITGCIGAGA TITGCIGAGA ACCAGACICC GAACACATTA GCAGACICC GAACACATA GCAGTANAGGA GCACITCAGA AGTAGAGAAC TCATITATA GCAGAACACA TAGGTGAGTA TAGGTGAGTA ACACACACACACACACACACACACACACACACACACA	AGUITMETT GGAATTITTC TGTTTGCATT PATGCATCA ATTGGGAAA GCAAGGAAGG TAAACTAGGC TACCACTGAT AAATAGATCA GTATGATCA AAATAGATCA AAATAAGATC TACGATGATATACA AAATAAGATC ACGGTGTATA AAATAAGATC CCGGGAACCAGA AAATATATTCT AAAGAGACT AAAGAGACT AAAGAGACT ACACCTGTGA ACCCTGTGA ACCCTGTGAAATAC ACGTGAAATAC ACGTGAAATAC ACGTTGAAAGA ACTCCATAAAA ACTCCATAAAA	TACGATICCC TAAGGATAGA TTCATGCTGC AGATGTCTT TAAAGCAGCA ACATTGATAGC AGATGCTGTG AGACCTGGTG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGTGGG ACCATCTGTGGG ACATGCCTA ATTATCTGT ACTATATA TTAACTATTA CGGGGAACCC ATAGAGTGAT TAAGAGTGAT TAAGAGTGAT TAAGAGTGAT TAAGAGTATA ACTCAGGAACCC ATAGAGTAACA TTAGGTATAA TTAGGTATAA TCCAATATA ACTCAGGAAC TGCAACTAC TGAGACACC GAACCTGGGA AAAAGTCTA GGGAACACC GAACCTGGGA AAAAGTCTA GGGAACACC GAACCTGGGA	TGGGGNAGG AGATAATACT CAGTATTAAT TCCCTTTAGC CATTCAGACA TGAGGCCAGG AGAGAATAA ACAGAGATAA TATGGAAGGC GGAGACATAA TATGGAAGGC CAGAGTCTGA TTTGTTGGGG AAACAGTAGT TATTATAAAA CCTGATTGTG AAACAGAAGA TCTGTGTGGA CTTTAATATAC TGATTCAAC TGATTCAAC TGATTCAAC TGATTCAAC TGATTAAAA TCAATTAAC TGATTCAAC TCACTTTAACT TACAGGATGA TGATTCAAC TCACTTAACT TACAGGATGAAC TCACTTAACT TACAGGATGAAC TCACTTAACT TACAGGATGAAT TACAGGATGAAC TCACTTAACT TACAGGATGAAC TCACTTAACT TACAGGATGAAT TACAGGATGAAC TCACTTAACT TACAGGATGAAC TCACTTAACT TACAGGATGATAC TCACTTAACT TACAGGATGATAC TCACTTAACT TACAGGATGATAC TCACTTAACT TCACTTAAC	ATAGATAGA AAGGGAGAAG GCAGCACTCT TTATTATTOG TTACTGAGTT TTATTGAGTT TGAGAGGTTT TTTTCATCTG GCATTGCAAG GCATTGCAAG GCAGCTAGAAC GAAATGGCTA AGGCTAGAAG CTATTTGTGAG CTATTGGATG CTATTGGATG TACAGCACGT AGGGTCAGGG AGGCTCAGGG AGAGTTTCCT ACAATAGTCA ACAATAGTCA ACAATAGTCA ACAATAGTCA CATCATCTG GCAGGACATTC GCAGGACATTCA TGCAGGGGTGG TGCAGGAGAAAAAAAA TAAGGGAAGAA	ATANGGATGA GCANTGTGAT CTTCTGCTAT AACAGTTTGG CAATATATGT GGTCTTAMA AACGTGAGGT ACCTTAAGAG CCGTAATGAA CTGAACAGTA GAGTCTTAGAG GTTTAGAGAG CTGAACAGTA GAGTCATGGG GATCATGGG GATCAGGAG AGTTAGGAG AATATATCAGG AATATATCAGA AATATCAGAA CCTGAACTTATAT AACATGAAA AATATTCAGG CATAGCATTT TATATCAGGA GCACAGCATT TATATCGGG AATATGCTGG CATAGCATCT AATATCGGA AATATGCTGG CATAGCATCT AATATCGGAG AATATCGAA AATATCGAAA AATATCAGAA AATATCAGAA AATATCAGAA AATATCAGAA AATATCAGAA AATATCAGAA CATTTAGCTGG CAACAGCATT AATATCCGAA AATATCGAA AATATCAGAA AATATCAGAA AATATCAGAA AATATCAGAA CATTTAGCTGG AACTTAGCTGA AATTTAGCTGA AACTTAGCTGA AACTTAGCTCAA	ACCTGGCTT  TCTGCAGCA  TANACTCTCA  ANTITIGANA  TGCCAGGTAC  GATANATGT  GATGAGTAGG  ATTICAGGAG  ATTICAGGAG  ATTICAGGAG  ATTICAGGAG  ATTICAGGAG  ATTICAGGAG  ATTICAGGAG  GCAGTACAAA  GACTTIGGT  TTACCTIGCA  TTACGTIGGA  GATAATGGAT  TTGGANAGAT  CTGGAGAGTG  GAGTTAATCA  AGATCAACAA  AGATCAACAA  ATTICAGAGT  ACCTGGAGAGT  ACTTGACAG  ATTITAGGAGT  ATTITAGGAGT  ATTITAGGAGT  ATTITAGGAGT  ACTTGACAT  ATTITAGGAGT  ACTTGACAT  ACTTGCACT  CATTCTCCGTT  CATTCTCCGTT  CCTTGGTCC  CCTTGGTCC  CCTTGGTCC  CCTTGGTCC  CCTTGCTTC  CCTTGCTCC  CATTCCCGTTC  CCTTGCTTC  CCTTGCTTC  CATTCCCGTTC  CCTTGCTTC  CCTTGCTTC  CCTTGCTTC  CATTCCCGTTC  CCTTGCTTC  CCTTGCTTC  CCTTGCTC  CATTCCCGTTC  CCTTGCTTC  CCTTGCTTC  CATTCCCGTTC  CCTTGCTTC  CCTTCCCTTC	TEGAGCETEGE TTGTAGCEAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAGT CGCAGGGCGC AAAAGGCGGG TTGAGTTAC AAACTGCATA AACCCTTAG AAACTGCATA AACCCTTAG CCTAGGAAGTAC ATTCAAATTAC ATTCAAATTAC ATTCAAATTAC ATTCAAATTAC CTAGGAAGT AAAAAGAGAT CTTAAGGTCGAT TTCACAGTCG TTCACGTCGAT TCCACGGCCAG CCACGGCCAGT CACCCTGGGC CACCCTGGC CACCCTGGGC CACCCTGGC CACCCTGGGC CACCCTGGC CACCCTGGGC CACCCTGGC CACCCTGGGC CACCCTGGC CACCCTGGC CACCCTGGC CACCCTGGC CACCCTGGGC CACCCTGGC CACCCTGGGC CACCCTGGC CACCCTGGC CACCCTGGC CACCCTGGC CACCCTGGC CACCCTGGGC CACCCTGC CACCCTGGC CACCCTGC CACCCTCC CACCC	GAMATATGA AGAGAGTAT TCTGTAMAG TACTAGGGGT TACTAGGGGT CTTAMAGGT TTCCATATGG ATGGAGGTAT CCTACAAGA AGACAACA TCAGAGGTAT TATGACCAC TATGAGCAC TATGAGCAC TATGAGCAC TATGAGCAC TCAGGAGCAC TCAGGAGCAC AGAGCATAAA ATATAAAGG TGGATGGTAT TTTAGACAC AACAGCATAA AACAGCTAC AACAGCTAC AGAACTAG TCAACCAC TCTACTGAA TCAACCACC TCTACTGAT AGAACTAG TCAACCACC TCTACTGAT AGAACTAG TCAACCACC TCAACGAGCAC AGAAATATG CACAACAC TCAACAGCAC	1411 1511 1711 1811 2011 2211 2211 2211 2311 2411 2511 2611 2711 3011 3011 3011 3011 3011 3011 4011 40
CCOTTGAGGA CGGTGAGGA CGGTGAGGA CAGTTTAGGTA ATAGGGAAAA CTAGGCAAA CTAGCACAA CTAGCACAA CTAGCACAA CTAGCACAA ATATTAGGA AGATTATAGG AGATTATAGGA AGATTATAGGA AGATTATAGGA AGATTATCTTCC CTCCAACACA TACTGACTA TACTTGTCA TACAGCAGTA ACAGCAGTTA ACAGCAGTTA ACAGCAGTTA TACAGCAGTTA TACAGTCAAC TACAGCAGTTA ACAGCAGTTA ACAGCAGCA CCTTTCAGCA ACTTCAGCAG ACTTCAGGAG ACACCAAC ACTTCAGGAG ACACCAAC ACTTCAGGAG ACACCAAC ACTTCAGGAG ACACCTAGAG ACACCTAGCA ACTTCAGAG ACACCTAGAG ACACCTAG	ACCIONADA ACCIONADA TTGCTGAGA TTGCTGAGA TTGCTAAAGTG ACCAGACTCC GAACACATTA GCAGACAGG GCATTTATGCTGGAGA GCAGACAGG GCAGTAGAGGAG CTAAGGAAG CTAAGGAAG CTAGGAGAGCAT TAGGTGAGTA CCTGCCAGTA ACACCCTGAGA ACACCTGAGAGCAT TAGATGAGTA GCAGCAATACCAG GATCATCCCTTG GATCCTGCCAGA ACGACCAATT ACGAGGAGAGG GCTCCCTTG GAGAGGGC AGTAGAGGG AGTTGAAGAG CTGCCAATACCAG AGTTGAAGAG CTGCCAATACCAG AGTTGAAGAG CTGCCAATACCAG AGTTGAAGAG CTGCCAATACCTGC AAAAAAAAAA	AGGITAGTA  TATGCATTA  TATGCATCA  TATGCATCA  ATGGGGAAA  GACAAGAC  TACCACTGAT  TAGCATTA  TAGCATATACA  TAGCATATACA  TAGCATATACA  TAGCATATACA  TAGCATATACA  AGGAACCA  GAGAACCA  TAGCATGAGAC  TAGCATGAGAC  TAGCATGAGAC  TAGCATCAC  TAGCATCAC  TAGCATCAC  TAGCATGAGAC  TAGCATGCAC  AGGAGATACCAC  AGGAGATACCAC  AGGAGATACCAC  AGGAGATACCAC  AGGAGATACCAC  AGGAGATACAC  ACTCCATAAA  AGGAGAGACAC	TACGATICECE TAAGGATAGA TICATGCTGC AGATGTCCT TAAAGCAGCA CAATTGATAG AGAGCTGGTC AGAGCTGGTC AGAGCTGGTC AGAGCTGGTC AGAGCTGGTC ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTATAA ATTOTCTGT GCTCAGTATA ATTOTCTGT GCTCAGTAAA ATTAGATATT GGAAAATTTA GGAAAATTTA GGAAAATTTA GGAAAATTTA ACCTCAGGAA ACCTCAGGAA TGGAACACA GAGCATCAA TGGTGGCTCA TAGATTTAACAT TTAGATTTAACAT TTAGATTTAACAT TTAGATTTAACAT TTAGATTTAACAA TTAGATTATAT ACCTCAGGAAC ACCTCAGAT ACCTCAGGAAC GGGAACACC GTGAACCAC GTGAACCAC GTGAACCAC GTGAACACCC GGGAACTAGC GTGAACAGCC GGGAACTAGC GTGAACAGCC GGGAACTAGC GTGAACAGCC GGGAACTAGC TCCAGGCAACA TCCAGGCAACAC TCCAGCCAC TCCACC TCCAGCCAC TCCAGCCAC TCCAGCCAC TCCAGCCAC TCCACC TCC	TGGGGNANGG AGATANATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGACTCA ATCAGATAAT CGATGCCCAG GGAGACATAA TATGGAGGCCAG GGAGACATAA TATGGAGGC TTTTTGTTGGGA AACAGTAGT TTTTTTTTTGTTGACACACATAAT CCTCATTGAAA CCTCATTAAAA CCTCATTAAAA TATTATAAAA CCTCATTAACT TGATGCAAATAGT TATTATAACT TGATTCAAAC TGATTCAAAC TGATTCAAAC TGATTCAAAC TGATTTAAAA TGCAGATAAT TATGGGGGG GGGGTAAAT TATGGGGGG GGGGTAAAT TATGGGGGG GGGGTAAAT GGGGGTAAAT GGGGGTAAAT GGCTTAAAT GGCTTAAAT GGCTTAAAT GGCTTAAAT GGCTTAAAT GGCTTAAAT GGCTTAAAT GGCTTAAAT GGCTTAAAT CCAATGAAAA CCAAATGAAAA	ATAGATAGA AAGGGAGAAG GCAGCACTCT TTATTATTOC TTACTGAGTT TGAGAGGTTT TGAGAGGTTT TTATCACTG GCTTTGCAAG GCATCATTA AGGGTAGAAC CAAATAGGCA AAGGCTAGAC TACCAGCACT TACAGCACT AAGGCCATGG AGGGCCATGG AGGGCCATGG AGAGTTCCT ACAATAGTCA AAGTAGGCC AATAGTCA AAGTAGGCC TACTGTTAAA CTATTTCTGT GCACATCACT CAATAGTCA AAGTAGGCC TACTGTTAAA CTATTTCTGT GCACATCACT CAACTATTA AAGCCATTTT AAGCCATCTC TGCAGGGTGG CCACCACTC TGCAGGTGG CCAGCACTT AAATACAAA GCAGTCACAC TACTGTTAAA TACAAAAAAAA TAAGGGAAGAC CAGGAAAGTT CAGAAAGTACAA TAAGGGAAAGAC CAGGAAAGTT CAGAAAATACAAA TAAGGGAAAGAA CAGAAAAAAA TAAGGGAAAGAC CAGGAAAGTT CCTGTTTTTA	ATAAGGATGA GCAATGTGAT CTTCTGCTATI AACAGTTTGC ACACTACGG CAATATAGG GGACTTAAAA AACGTGAGGA CCGTAATGAA CTGAACAGTA GAGTTTGGG GATTTAGTATG GATCAGGAG GTGTGGGCAT CCGGACTAGA TGGTTCACTC AGAGGAACAGTA AACATGAAA CTGATCAGAT TATATCAGA CCTAATTACA CATAGCATT TATATCAGA CATAGCAGA AAGTTTAGTT AACATTAAAA AACATTAAAA AACATTAAAA CATATATAT	ACCTGGCTT TCTGCAGGCA TAMACTCTCA MATTTTGAA TGGCAGGTAC GATAMATGT GATGAGTAGG ATTTGGACAA ATTTGGACAA TTCGAGAGAC ATTGGACAA TAGATGGTG GCAGTACAAA GACTTTGTG TTACCTTGCA GATAMATGGT TTACCTTGCA GATAMATGAT TTACGAGAGTA CTGGAGAGTA TAGATAGGAT TAGATAGGAT TAGATAGGAT TAGATAGGAT TAGATAGGAT TAGATAGGAT TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TTTAGAGGT AAAATGTGAT ACTTGACCAA GCAACGAGCT GCAACGAGCT AAAGGAACT CATTCCCGTT CCCTTGGTCC CATTCCCGTT CCCTTGGTCC CATTCCCGTT	TEGAGCCTEG TTGAGCCAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG CCACTAGGTA CGTTCTCTT TTTTGTATGG AAAATGCAAT GCGAGCGC AAAAGAGGGG TTGATCTAT AGCATGCATA AGCATGGATA AGCATGGATA AGCATGGATA AGCATGGATA ATTCAAGTGC TTAGAATTAAG TTAAGGATGA ATTCAAGTGC TTAGAATTAAG TTAAGGATGA AAAAGAGAT TTAAAGGATGA AAAAGAGAT TTAAAGGATGC TTAGAAATTAAG CCACAGATCT TCAAAGGCC GAAGGAAAT ACCTTGGAT GCGTGCCTGT CCACAGTCC GAAGGAAAT TCACAGTCC GAAGGAAAT TCACAGTCC GAAGGAAAT TCACAGTCC GAAGGAAAT TCACAGTCC GAAGGAAAT TCACAGTCC GAAGGAAAT TCACAGTCC CACAGTCC CACAGTC CACAGT	GAMTANTGA AGAGANTTAG ALAGGAGTTT TCTGTAMAG TACTAGGGT TTAGAGGATC CTITAMATGT TTTCGATATGG ATGGAGGTAT TGTGCAGATA AGTTCAMATG TATGGAGGATG TATGGAGGATG TATGGACACAGT TCAGGAGGAG TCAGCACAGA ATGTGAATACC AGGATGGATAACAACAAC ACACCACCAC AGGATGGTATTTGC AGAACATGG TGATATTTGC GAGAAATATGG CACAACAAC AGAACATGGA TGATATTTGC GAGAAATATGG CACAACAAC CACAACAAC AGAAATATGG CACAACAAC CACAACAAC CACAACAAC CACAACAAC	1411 1511 1711 1911 2011 2211 2311 2411 2511 2511 2711 2911 3011 3111 3211 3311 3411 3511 4011 4111 4511 4511 4611 4712 4811
CCOTTGAGGA CGGTGAGGA CGGTGAGGA AGGGATAA ATAGGGAAAA CGAAGTATAAG ACTATGACAA CGAAGTATAAG ATATGACAA CGAAGTATAAG ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA ATATTAGGA AGAAGATTA TGATACAGTA TGATACAGTA ACAGGAGTTA TGATACAGTA ACAGGAGTTA TGATACAGTA ACAGGAGTTA TGATACAGTA ACAGGAGTTA AGGGGGTTT TGATAGGA AGATGATTA AGGGGGTTT TGATAGGA AGATGATTA AGGGGGTTT TGCTAGGA AGTGACAGGA ACTTGACAGGA ACTTGACAGGA ACTTGACAGGA ACTTGACAGGA ACTTGGGGG ACTTGACAGGA ACTTGGGGGG TGCTTTTCTC ATTGCTAGGA CCACAACAG ACTTGCTAGGA CCACACAGG ACTTGCTAGGA CCACACAGGG ACTTCCTAGGA CCACACAGGGGGT ACTTGCTAGGA CCACACAGGGGGGGGGG	ACCIMANANA ACGGGCCAGG TITGCIGAGA TITGCIGAGA ACCAGACICC GAACACATTA GCGCATTITAT GTGAGACAGG GCATTCCIGA AGTAGAGAC CTAAGAGAC TCATITATAA GCAGAACCAT TAGGIGAGTA CCTGGCAGTA AAAAGCTGAG AACAACCA TTAAGAGCT TTAAGAGCT CATAAACTGA GCACAACTC CATAACCAG GCTCTCCITG GATCCIGCA AGGICACAGA CTCTCTITATC AAGTACAGAG GCACCAGTA CTGGCCACA CTGGGCACA CTGGCCACA CTGGCCACA CTGGCCACA CTGGCCACA CTGGGCACA CTGGCCACA CTGGCCACA CTGGCCTTGAACA CCTGGCTTGAACA CCTGGCCTGAACA CTGGCCTTGAACA CCTGGCCTGAACA	AGGITAGTA TATGCATTA TATGCATCA TATGCATCA TATGCATCA TATGCATCA TATGCATCA TATGCATCA TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCATTGAT TACCATTGAT TACCATTGAT TACCATTGAT TACCATCGAT TACCATCAT TACCATCGAT TACCATCAT TACCATC	TAGGATAGA TTAAGGATAGA AGATGTCCTT TAAAGCAGCA AGATGTAGTGC CAATTGATAGA AGAGCTGGTC AGCATCTGTG ACCATCTGGG GTCTGACTTG ACATTGCTATAAAATATATATATATATATATATATATATA	TGGGAMAGG AGATAMATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGACATAA CAMAGACTCA ATCAGATAATA TGCATGTCCAG GGAGCCTAGA TCTGTGGGA CAGAGTCTGA TCTTGTGGGA AMACAGTAGT TTTTGTTGAC TGTCCCTAC TGTCCTAC TGTCCCTAC TGTCCCTAC TGCTCTACAT TACTGTGGG GGGTTACTTA TACTGTGGGG TCCATTAAAT CCAATGAAAA CCAATGAAAA CCAATGAAAA CCAATGAAAA CCAATGCGGG TCCATCCTA ACAATGCCC ACCAATGCCC ACCAATGCCC ACCAATGCCC ACCAATGCCC ACCAATGCCC ACCAATGCCC ACCAATGCCC ACCAATGCCC ACCAATGCCC ACCAATGCCCC ACCAATGCCC ACCAATGCCCC ACCAATGCCC ACCAATCCC ACCAATGCCC ACCA	ATAGATAGA  AAGGGAGAAG  GCAGCACTCT  TTATTATTCC  TTATTATTCC  TTATTGAGTT  TGAGAGGTTT  TTTTCCATGAGT  ATCTGCAAG  GCAGTAGACA  AGGGTAGACA  AGGGTAGACA  ATCCTGAATG  CTATTTGTGAG  ATCCTGAATG  CTATTTGTGAG  ATCCTGAATA  AGGCACACGT  AGGGTAGACA  AGGGCTAGGC  AAGTCATTCCT  ACAATAGCA  AAGTAGTCAAT  AAGTCATTCCT  ACAATAGCCA  CTATTTCAG  CCACCACTGT  AAACTATTCAA  AGCAGGTGG  CCCAGCACTT  AAACTATTTAAA  CCACTATTTAAA  CAACTATTTAAA  CAACTATTTAAAA  CAACTATTTAAAA  CAACTATTAAAAAAAA	ATAAGGATGA GCAATGTGAT ACAGTTTCG ATATCTGCTAT ACAGTTTCG ACACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTTAGGAG TITTAGTATG GATCATGG TITTAGTATG GATCAGGAG TGGTCGGCAT TGGTCAGCAG TGGTCAGCAG TGGTCAGCAT ACATAGAAG TGGTCACCC AAAAAGGGAG AAAAAGGGAG AATATTCAAG CATAGCATT TATACCAGG ATAGCAGT CAATTTCAT TATACCAGG ATAGCAGT CAATTTCAT CATTTCAT TATACCAGG ATAGCAGT CAATTTCAT CATTTCAT TATACCAGG ATAGCAGT ACCACCAC AAGTTTAGTA ACATTCAGT ACATTCAGT TCAGAGGCCC AATTTAGTA AGAATCGCCA AAGTTTAGTT ATATACAG AGAAATCGCCA AAGTTTAGTT ATATACAG AGAAATCGCCA AAGTTTAGTT ATATACAG AGAAATCGCCA AAGTTTAGTT ATATACAG AGAAATCGCTA	ACCTGGCTA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA GATANATGT GATGAGTAGG ATTCGGCAG ATTCGGCAG ATTCGGCAG ATTCGGCAGAG GCAGTACAAA GATCACCTGCA GATANATGGT TTACCTTGCA GATANATGGAT TTACCTTGCA GATTATGGAATTG CTGGGAGTTG CTGGGGAGTAG TANACTGAT TANACTGAT TANACTGAT ACTTGGAATTG ACTTGACAAA TANATGGAT TATTGACAAA TATTGACAAA TATTGACATA ACTTGACCAT ACTTGACCAT ACTTGACCAT CCATTGGCCC CCATGGCCC CATTGCCCTC CCATTGCCCT CCCTTGGTCC CATTCACCAC CATTAAGCAA CATTAACTAA CATTACTACTAA CATTCCTCTAC CATTACTACTAA CATTCCTCTAC CATTACTACTAA C	TEGAGCCTEG TTGAGCCAG TTGGCTTCTA TTAMACAGT TTGTCTAGGG CCACTAGGTA COTTCTCTT TTTTGTATGG CACTAGGTA AMATGCAT TGGTCAGG TAGATGTAG AMATGCAT TGGTCAGG TAGATGTAG AMACGGGG TTGATGTAGAT ACCATGGATA ACCATGGATA ATCAGGATCA TTAGAATTAA TAAGGATCA TTCACAGTCC GAAGGAATTAAG CCACAGATCT TCCACAGCCCAG TTCACAGTCC GAAGGAATTAAG CCACAGATCT TCCACAGCCCAGATCT TCCACAGTCCAGATCT TCCACAGTCC GCAGGGCCAGATCT TCCACAGTCC TTCACAGTCC TTCACAGTCC TTCACAGTCC TTCACAGTCC TTCACAGTCC TTCACAGTCC TCACAGCCAGATCT TCTACAGTTCATCATCATCATCATCATCATCATCATCATCATCACTCCCCCC	GAMTANTGA AGATANTAGA AAAGGAGTTT TCTGTAMAAT TACTAGGGGT TTACAGGGTT TTCGAGAGGATC TTACAGGGTAT TCGAGCAGA AGTCAAAT TCAGCACAGA AGTCAAAT TCAGCACAGA ATTCAAAT TCAGCACAGA ATTCAAAT TCAGCACAGA ATTCAAAT TCAGCACAGA ATTCAAAT AGACTACA AACAGCAT AAACCAGCAT AAACCAGCAT AAACCAGCAT TCAGCAGCT TCAGCAACCT TCAG	1411 1511 1711 1811 2011 22111 22111 22111 2311 2411 2611 2711 2611 2711 3011 3111 3211 3311 3411 3511 3713 4011 4211 4311 4311 4511 4611 4711 4811 4811 5111
CCOTTGAGGA CGGTGAGTGA ATAGGGATAA ATAGGGATAA ATAGGGAAAA CGAAGTATAAG ACTATGACAA ACTATGACAA ACTATGACAA ACTATGACAA ACTATGACAA ACTATGACAA ACTATGACAA ACTATGACTA TGTATCTTCC CTCCAACACA TATCTGACTA TGAAAAGGA ACAGGAGTTA TGCTGAGCA ACTTGACGA ACTTGACGA ACTTGACGA ACTTGACGA ACTTGAGGA ACTTGAGA ACTTGAGAG ACTTGAGAG ACTTGAGAG ACTTGAGAG ACTTGAGAG ACACAAACA AUTAGAGTGT ACACAAACA AUTAGAGTGT ACACAAACA AUTAGAGTGT ACACAAACA AUTAGAGTGT ACACAAACA AUTAGAGGTGT ACACAACAACA AUTAGAGGTGT ACACAACAACA AUTAGAGGTGT ACACAACAACA AUTAGAGGTGT ACACAAACA AUTAGAGGTGT ACACAACAACA AUTAGAGGTGT ACACAACAACA AUTAGAGGTGT ACACAACAACA AUTAGAGGTGT ACACAACAACA AUTAGAGGTGT ACACAACAACA AUTAGAGGTGT ACACAACAACAACA AUTAGAGGTGT ACACAACAACAACA AUTAGAGGTGT ACACAAACAAACA AUTAGAGGTGT ACACAAACAACAAACA AUTAGAGAGAACAAACAA AUTAGAGAGAACAAACAAACAAAACAAACAAACAAAACAA	ACGGGCCAGG TITGCITGAGA TITGCITGAGA ACCAGACTCC GAACACATTA GGGATTTTAT GGGAGACAGG GCATTTATAGA GCAGATCCTGA AGTAGAGAGC CTAAGAGAAC TCATTTATAA GCAGAACCAT TAGGTGAGACA TAGGTGAGA AAAAGCTGAG GAGCAACACACACACACACACACACACACACACACA	AGUITMETT GGAATTITTC TGTTTGCATT PATGCATTCAT ATGGGAAA CTCTTTGATC ATTGGGGAAA GCAATGGGC TACCACTGAT GACTATTGCA TTGATCATTGCA TTGATCATTGCA TTGATCATTGCA TTGATGATTA AAATAAGATC TTGGTTGTAT AAATAAGATC CTGGTGGATCATA AAGAGAATTATTCT AAAGAGAATTATTCT AAAGAGATTCAC TTGATGATGAT TTCATGATTTTCATGAT TTCATGATTTCATGAT TTCATGATTCATATTTCATGATTACACTTGAA AGGATTCACA AGGATGCATC AGATGCATGAAA AGGATTGAAGAA AGGATTGAAGAA AGGATTGAAGAA AAGGAATCA AATAGCATAAA GAGTTGAAGAA AAGGAATCACA AAGGAACAAC ACCTAAAAA AAGGAACAAC ACCTAAAAGCAA AACACTGAAA AACACTACA	TACGATICCE TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAGGATAGA ATCATCGTG AGAGCTGGTC AGAGCTGGTGT AGAGCTGGTGA ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCATTAAATAT GGGAGAACCC ATAGCTCAT TIAGCTATAT GAGAGATAT GAGAGATAT GAGAGATAT GAGAGATAT TAAGGTTAAA ACTCGAGAAC TCCATTAACAT GCAATTTAG GGAACCC GAACCTGGGA AAAAGTCTA GTGAACACC GGAACTACT TCCATCCTT TATAGCTACT TATAGCTTAT TTAGCTCAT TTAGCTCAT TTAGCTCAT TTAGCTCAT TTAGCTCAT TTAGCTCAT TTAGCTCAT TTAGCTCAT TTAGCTCAT TTAGCTTAA	TGGGGNAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGACTCA ATCAGATAAT CAAAGACTCA ATCAGATAAT TATGTGAGGG CAGGCCTAGA TATGGAAGGG CAGAGTCTAA AACAGTAGT TTTTTTGTGGA CTGATTGTG AAACAGTAGT TATTATAAAA CCTGATTGTG AAACAGTAGT TATTATAAAA CTTATTATAAC TCAATTAACT TCAATTAACT TCAATTAACT TACAGCATAAT TACAGCATAAT GATGGAGGTT CACCTATAAT CACCTATAAT CACCTATAAT CACCTATAAT CACATAAAT CCAAAATAACT CAAAAACCCC TTCTTGAGTT AAACAGAGGT TACAAATACCT CAAAAACCCC TTCTTGAGT AACAAACCCC TACAATACCT AACAATACCCA TACAATACCCA TACAATACCCA TACATACCTA TACACCTATA	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTOC TTACTGAGTT TGAGAGGTTT TGAGAGGTTT TGAGAGGTTT TGAGAGGATCA AGGCTAGAAC AGGCTAGAAC AGGCTAGAAC ATACTGAGTA AGGCTAGAAC AGGCTAGAAC AGGCTAGAAC AGGCTAGAC AGGGCCATG AGAGTATAC AAGTAGGCC TACTGTTAAA CTATTTCTGAG GCAGTATAC AAGTAGCC AAGTAGCAC TACTGTTAAA CTATTTCTGT GCAGCACTT AAATACTA AGGCACTT TGATTACTAC CAATTATTA AGGCATATAT AAGCAATTT AAAGCAATTT AAAGCAATTT AAAGCAATTT AAAGCAATTT AAAGCAATTT AAAGCAATTT AAAGCAATTT AAAGCAATTT AAAGCAATTT AAATACAAA CCAGTGAGCC CCCCAGCACTT AAATACAAA CAGGAGAGC CCCTGTTATTAG GGTTGTATTG	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG CAATATAGGT GGTCTTAAAA AACAGTTAGGA CCGCAATGAGAG CCGCAATGAAC CTTAAGGAG CTTAAGAAGTA GATCATGGG CTTTTAGTATG GATCAGGGA CTGGGCAT AGAACAGTA AACATGAAG AAAAGGGAG AAAAAGGAG AAAAAGGAG AAAATGGAA CCTAATTAAA AACATGAAG AATATCAGAG CCTAATTTAC CATATCAGAG CATATCAGA CATATCAGAG CATATCAGA CATATCAGAG CATATCAGA CATATCAGA CATATCAGA CATATCAGA CATATCAGA CATATCAGA AAGATATAAGA AAGTTAAGAA AAGTTAATAAGA AAGTTAATAAGA AAGTTAATAAGA AAGTTAATAAGAA AAGTTAATAAGAAAAAATAAGAAAAATATGAAAAAAAA	ACCTGGCTT ACCTGGCA TAMACTCTCA AMTTTGANA TOGCAGGTAC GATANATGT GATGAGTAGG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG ATTICAGGAG GCAGTACAAA GACTTTGTGT TTACCTTGGCA GATANATGGAT CTGGAGAGTA CTGGAGAGTA CTGGAGAGTA CTGGAGAGTA TAGATCAACAA TAGATCAACAA TAGATCAACAA TAGATCAACAA TAGATCAACAA TAGATCAACAA TAGATCAACAA TAGAACAAG ATTITACCA GCAAACAGG GCATGGGGG GCATGGGGG CACTGCAGCC CACTGCAGCC TACAACAGCA TAGAACAACAAC AGATCAACACAC CACTGCAGCC CACTGCACC CACTACC CACTGCACC CACTACACC CACTGCACC CACTCACC CACTGCACC CACTGCACC CACTCACC CACTCACC CACTCACC CACTCACC C	TEGAGCCTEG TTGAGCCAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA AGATTCTCT TTTTGTATGG AAAATGCAGT GGAGGGGG AAAAGAGGGG AAAAGAGGGG TTGAGTTAG AACCCTTAG AAACTGAAT AGCATGGAT ATCAAGTTAC TTAGAAATTAAAATTAAGATGC TTAGAAATTAACATGGA ATCATGAAATTAAGATGC TTAGAAATTAA TAGAAATTA ACCATGGAAT ATCATGAAATTAAGAAGAAGAAATTAAGATTAC GTTATTAAGGTC GTATGAAATTAAGAAGAAGAAATTAAGAAGAAATTAAGATTC GTTATTTCAAC TTCACAGGCCA TTCACAGGCCA TTCACAGGCCA TTCACAGGCCA TTCACAGGCCA TTCACAGGCCA TTCACAGGCCA TTCACAGGCCA TTCACAGGCCA TTCACAGGCAATTACGATG CAGCCTGGGC CAGCCTGGGC CAGCCTGGGC TATTAGGATG TTTTCATCATT	GAMTANTGA AGAGANTAGA ALAGGATTT TCTGTAMAGA TACTAGGGGT TTACATGGGT TTCCATATGG ATGGAGGTAT TCTGACATG AGTTCAAATG TATGGACGATG TATGGACGATG TATGGACCAGA AGTTCAAATG TATGGACCAGA AGTTCAAATG TCAGGAGACAGT TCAGGAGACAGT TCAGGAGACAGT TCAGGAGATGAA ATGATATAGG TCAGCACATGA ACCAGCATAA AGCAGCATGA ACCAGCATGA TCTACTGACACAC CCTCTACTGA ACCAGCATAA TGCAGCATGT TCAGGAGGT TCAGGAGAT TCAGGAGAC TCAGAGACAC TCAGAGACAC TTACAGGAA TCCAGACAC TCAGAGACAC TCAGAGACAC TCAGAGACAC TCAGAGACAC TCAGAGACAC TCAGAGACAC TCAGAACAC TCAGAGACAC TCAGAACAC TCAGAAC	1411 1511 1711 1811 2011 22111 22111 22111 2311 2411 2511 2611 2711 3011 3111 3211 3311 3511 3511 4011 4111 4411 4511 4611 4611 4611 4611 5011 5011
CCOTTGAGGA CGGTGAGGA CGGTGAGGA ATAGAGAAAG ACTATGAGAAAG ACTATGAGAAAG ACTATGAGAGA ACTATGAGAGA ACTATGAGAGA ACTATGAGAGA ACTATGAGGAGA ATATTAAGGA ACTATGACTA TGATCTTCC CTCCAACCA TATTTAGGA TATTTAGGA ACAGAGGTTA ACAGAGGTATA ACAGAGGTTA ACAGAGGTATA ACAGAGGTATA ACAGAGGTATA ACAGAGGTATA ACAGAGGTATA ACAGAGGTATA ACAGAGGTATA ACAGAGGTATA ACAGAGGTATA ACAGACCAGC ACTTCGAGA CCACACCAGC ACTTCGAGA CCACAACCAAC ACTTCGTAGA CCACAACCAAC ACTTCGTAGA ACACAAACA ACACAAACA ACACAAACA ACACAAACA ACAAACAAACA ACACAAACA ACACAACA	ACCIMANDA AGGGGCAGG TITGCIGAGA TITGCIGAGA ACCAGACTCC GAACACATTA GGGATITTAT GTGAGACAGG GCATTATGGTGGAGA ACATANAGGG CATANAGGA CTAAGAGAAC TCATTTATAA GCAGAACCATTA TAGGTGGGAGA ACAGAGCAGA CATAAGAGAG CATAAGAGAC CATAAGAGAC TCATTTATAA GCAGAACCAGT TAAAGAGTGAG GAGCAAGCG CATAAAGCTGAC AAGGTCACAG GCTCTCCTTG AAGGTCACAG CTGTGCAGAA ACGTGAGAC CTGGCCAACA CTGAGGCAGAC CTGGCCAACA CTGAGGCAGC AAAAAAAAAA	AGGITMETT GGAATTITTC GGTTTGCATT PATGCATCAA CTCTTTGATC AATGGGAAA GCAAAGAAGG TACCACTGAT GACTATGCA GTATGACTTA AATAAGATG TACATTATACA CTATGACTTA AAATAAGATG TACATATACA CTGGTTGTAT AAATAAGATG AGGTGAGACT CAGGTGACT AAAGAGACT AAAGAGACT AAAGAGACT GAGGTCAGA TCCATGAGAC AGCCTGGAA AGCATCCAT AAGGACT AGAGTCCAC AGCCTGGAA AGCATGAGT AGAGTCCAC AGCCTGGAA AGCATGAGT AGATGCAC AGCCTGGAA AGCATGAATG AGCATGAATG AGTTGAAGAG ACTCATAAA AAGGAACAC AAGGAACAC AAGGAACAC AAGGAACAC AACCTGTAG AAGCAACAC AAGGAACAC AAGGAACAC AAGGAACAC AACCTGTAG AACCTTAAATGG AACCTGTAA	TAGGATAGA TTCATGCTGC AGATGTCTT TAAAGATAGA AGATGTCTT TAAAGCAGCA AGATGTATAGA AGAGCTGGTC AGAGCTGGTC AGACTGCTA ACATTGCG GCTGATTAA ACTTACTGTG GCTCAGTATAA ACTTACTGTG CCTCAGTAAA TCATTACTATTA GATAAATTAC GAAAATTAA TAGGTATAA TAGGTATAA TAGGTATAA ACTGGGAAACAC TAGAGTAAT ACTGAGGAAC CCCTAAACAA TTAGGTATAA ACTGGGAACAC GAACCTGGGA AAAAGTCTA GTGAACAC TGGAACACT TGGAACACT TGGAACACT TGGAACACT TCCAGTACA TCCAGCTAC TCC	TGGGAMAGG AGATAMATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCAGG AGAGAATAAT GGATGCCAGG GGAGACATAA TCAGGATATA TGCTGTGGGA TCTGTGGGA TCTGTGGGA TCTGTGGGA TCTGTGGGA TCTGTTGGA TATATATAAA TATTATAAAA CCTGATTGTG TAATTATAAA TATTATAAAA TATTATAAAA TATTAT	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTGC TTATTATTGC TTACTGGGTT TGTGCAGGTT TGTGCAGG GCATTGCAGG GCATTGCAGA AGGCTAGACA AGGCTAGACA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA CAATAGCTA AGGGCTACACA AGGGCTACACA AGGGCTACACA AGGATACACACTCT ACAATAGCCA CAATAGCCA CAATAGCCA CAATACCACTC TACCACTCC TACTCC TACCACTCC TACCACTC TACC	ATAAGGATGA GCAATGTATT ACAGTTTCG ATTCTGCTATI ACAGTTTCG ACACTATCG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTTAAGGAG CCTAATGAT CTGAACAGTA GAGTCTTCGG TTTAGTATG GATTAGAAGTGA ACTTAGGAG GTTTGGGCAT TGGTCACTC ACAGGACAGA TGGTTCACTC ACAGGACAGA TTCATTATA CCTAATTTAC CCTAATTTAC CCTAATTTAC CATTAGCAT TATATCAGG ACTAGCATT TATATCAGG ACGACCAT TATATCAGG ACGACCAT TATATCAGG ACGACCAT TATATCAGG ACGACCAT ACCTTAGTAT AACATTAGAG AGGATGCCCA AGGTTAGTAG ATGTTATAGA ACATTAGAG ATGTTATAGA ACATTAGAA ACATTAGAAA ACATTAGAAAA ACATTAGAAAA ACATTAGAAA ACATTAGAAA ACATTAGAAAA ACATTAGAAAA ACATTAGAAAA ACATTAGAAAA ACATTAGAAAAAAAAAA	ACCTGGCTT TCTGCAGCA TANACTCTCA ANTITTCAN TGGCAGGTAC GATANATGT GATAGGTAGG ATTCGAGGAG ATTCGAGAGA ATTCGAGAGA ATTCGAGAGA ATTCGAGAGA TANACTGG GCAGTACAAA GATTAGTGG TTACCTTGCA GATAATGGTA TTACCTTGCA GATAATGGAA TTAGAGAGAGT TTGGAGAGTA TCTGGAGAGTA TAATGGTAA TAATGGTAA TAATGTAT TATTGACAA TTAATGTAT TATTGACAA TTAATGTAT TATTGACAA TAATGTAT TATTGACAA TATTTTTTCC CATTGCACTC AAGGAACTT CATTCCGTTG CACTGCACTC AAGGAACTT CACTTCCGTTC CACTTCCTAG AGGTACATCA ATTCCCCTAG AGGTACATCA ATTCCCCTAG AGGTACATCA ATTCCCCTAG AGGTACATCA ATTCCCCTAG AGGTACATCA ATTCCCCTAG AGGTACATCA TTCCCCTAG AGGTACATCA TTCCCCTAG AGGTACATCA TTCCCCTAG AGGTACATCA TTCCCCTTAG AGGTACATCA TTCCCCTTAG AGGTACATCA TTCCCCTTAG AGGTACATCA TTCCCCTTAG AGGTACATCA TTCCCCTTAG AGGTACATCA TTCCCCTTAG TTCCCCCTTAG TTCCCCTTAG TTCCCCCTTAG TTCCCCTTAG TTCCCCCTTAG TTCCCCCCTTAG TTCCCCCTTAG TTCCCCCCTTAG TTCCCCCCTTAG TTCCCCCTTAG TTCCCCCCTTAG TTCCCCCCTTAG TTCCCCCCTTAG TTCCCCCCTTAG TTCCCCCTTAG TTCCCCCCTTAG TTCCCCCCTTAG TTCCCCCCCCCTTAG TTCCCCCCCTTAG TTCCCCCCCTTAG TTCCCCCCCCTTAG TTCCCCCCCCTTAG TTCCCCCCCTTAG TTCCCCCCCCTTAG TTCCCCCCCTTAG TTCCCCCCCCCC	TEGAGCCTEG TTGAGCCAG TTGGCTTCTA TTAMACAGT TTGTTCAGGG GCACTAGGTA COTTCTCT TTTTGTATGG GCACTAGGTA AMATGCAGT AGGCGCG AMAGAGGGG TTGATGTGAT AGCCCTTAG AMACCCTTAG AMACCCTTAG ACCCTTAG CCTAGCAAGT ACCATGGATA ATTCAAGTAC TTAGAATTAA TAAGGAAGCA TTAGAATTA AGCAGGCCCA TTAGAATTA AGCATGGATA AGCATGTAC TTAGAATTA TAAGGAAGCA TTAGAATTA TAAGGAAGCA TTAGAATTA TTCATGAT CCACGGCCCA TCCACGGCCCA TCCACGGCCA TCCACGGCCCA TCCACGGCCA TCCACGCCA TCCACGCCACA TCCACGCCA TCCACGCA TCCACGCCA TCCACCA TCCACCAC TCCACCA TCCACCA TCCACCAC TCCACCA TCCACCA TCCACCAC TCCACCA TCCACCAC TCCACCA TCCACCAC TCCACCAC TCCACCAC TCCACCAC TCCACCA	GAMTANTGA AGATANTAGA AAAGGAGTTT TCTGTANAAT TACTAGGGGT CTHAAATGT TTCCATATGG ATTGAGGTAT ATTGAGGTAT AGTGAGGTAT TATGAGGTAT TATGAGGTAT TATGAGGTAT TATGAGCACA AGTCANAATG TATGAGCACA TTATGAGCACA TTATGAGCACA TTATGAGCACA ATATATAGG TGAATGGTAT AGAGCATGA AGAGCTTGTG TATACACACA AGAGCTTGTG TCAACACACA AGAGCTTGTG TCAACACACA AGAGCTTGTG TCAACACACA AGAGCTTGTG TCAACACACA AGAACCTAG AGAAACTAGA AGAACCTAGA AGAACCTAGA AGAACCTAGA AGAACCTAGA AGAACCTAGA AGAACCTAGA TCAAGAGCAT TCAAGAGCAC GCAAAAGCAG GCATTTTACT TTATCAGACA ATATACACACAC TTATCAGACAC TCAAGAGCAA AGAGCACAC TCAAGAACACAC TCAAGAACACAC TCAAGAACACAC TCAAGAACACAC TCAAGAACACAC ATTATAAACCT TCAAAACCT TCAAGAACCT ATTATAAACCT TCAAAACCT TCAAAACCT TCAAAACCT TCAAAACCT TCAAAACCT TCAAACCT TCAAAACCT TCAAAACC TCAAAACC TCAAACCACAC TCAAACCACAC TCAAAACCACAC TCAAAACACAC TCAAAACACAC TCAAAACACAC TCAAAACACAC TCAAAACACAC TCAAAACACACAC	1411 1511 1711 1811 2011 22111 22111 22111 2311 2411 2611 2711 2611 2711 3011 3111 3211 3311 3411 3511 3713 4011 4211 4311 4311 4511 4611 4711 4811 4811 5111
CCOTTGAGGA CGGTGAGGA CGGTGAGGA AGAGTATAAG ATTAGGGAAAA CGAAGTATAAG ACTATGACAA CGAAGTATAAG ATTATGACAA AGAAGTATAAG ATTATGACAA AGAAGTATAA AGAAGACACA TAGTACAGCA TATTAGGA AGAATATTAGGA AGAAGTATA TAGTACAGTA TAGTACAGTA TAGTACAGTA ACAGGAGTTA ACAGGAGTA ACAGGAGTATA ACAGGAGTA ACAGGAGTATA ACAGGAGTA ACAGGAGGA ACTTGAGAG ACTTGAGAG ACTTGAGAG ACTTGAGAG ACTTGAGAGT TAGGAGGACA ACTTGAGAGT TAGGAGGACCA ACTTGAGAGT TAGGAGGACCA ACTTGAGACT TAGGAGCACA ACTTGAGACT TAGGAGCACA ACTTGAGACT TAGGAGCAC ACTTATATATG	ACCIONALE ACCIONALE ACCIONAL A	AGTIMETT GGAATITITC GGTATGATT TATGCATCA AATGGGAAA GCAAGAAGG TACCACTGAT CACTATTGATC AATGGGAAA GGAATAGACTA AGGATATAGAC GTATGACTTA AAATAAGATT AAATAAGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGAGGAC TATCATTATT ACACCTGTGA AAATTATTCT ACACCTGTGA AAATTATTCT ACACCTGTGA AAATTATTCT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT ACACCTGTGAA ACACTGTAGAC ACTTGAAATG ACTTGAACAC ACTTGAACAC AACACTGTAG	TAGGATAGA TTAAGGATAGA AGATGTCCT TAAAGCAGCA AGATGTCCT TAAAGCAGCA AGATGTAGTG AGAGCTGGTC AGAGCTGGTC AGAGCTGGTC ACACTGTGG GCTCATTIAA ATTATCTGTG GCTCAGTATAA ATTATCTGT GCTCAGTAAA TTAAGATTATATAT TAGGTCATT AGAGATATT AGAGATATT AGCATCAT ACCCACAGAT TAGAGTACT ACCCACAGAT TAGAGTACT ACCCACAGAT TAGAGTACT ACCCACAGAT TAGAGTACT ACCCACAGAT TTAGACATCA ACTGAGGAAC GAACCTGGGAA ACAGCTAA CTGAACAGCT GGAACAGCT TGAACAGCT TGCACAGCTAA TCCATTCCTT TATAGGTTACT TATAGGT	TGGGGNAAGG AGATAATACT AGATATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGACATCA ATCAGATAATACT GGATGCCCAG GGAGACATAA ATCAGATACA TTTTGTGGGA CAGAGTCTGA AAACAGTAGT TTTTGTTGAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTATTATAAA CCTGATTTATAC TCAATTAAAA TGCAAATAAC TCAATTAACT TACAGGAGGA CACTATAAT GGGTGTAGAT GGGTTAAATT GGGTGTAGAT GGGTTAAATT CCAATGAGA CATGAAAAA CCAATGAAAA CCAATGAAC TACATGCGG TACATACCTA CAGTGGTTT TACATATACCTA TACATTACCTA TACATTACCTA TACATTACTATT	ATAGATAGA  AAGGAGAAG  GCAGCACTCT  TTATTATTGC  TTATTATTGC  TTATTGGGTT  TGAGAGGTTT  TGTACTGG  GCTTGCAAG  GCAGTAGACA  AGGGTAGACA  AGGGTAGACA  ATCCTGAATG  CTATTTGTGAG  ATCCTGAATG  CTATTTGTGAG  ATCCTGAATA  AGCCACTGT  AGGGTAGACA  AGGGCTAGGC  AGGGTTGCT  AGAGTAGTCA  AGGGTAGACA  AGGGTAGACA  TATGATATAC  CAATTGTCA  AGGGTAGGC  ACATTGTTAAA  CAATTGTTAAA  CAATTGTTAAA  CAATTATTCGG  CCCAGCACTT  AAAGTAATACACA  AAGGGAGGG  CCCAGCACTT  AAAGTAATACACA  CAATTGTTAAA  CAATTATTCGG  CCCAGCACTT  AAAGTAATACACA  TAAGGGAAGA  CAGGAAGATCCTTTATTAA  GCAGTGTGGC  CCTGTTATTA  ATTATTCTGG  GGTTGTATTCA  GTTGCACAAA  ATTATTCTGG  GGTTGTATTCA  ATTATTCTGG  GGTTGTATTCA  ATTATTCTGG  GGTTGTATTCA  ATTATTCTGA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TTATTATTCTA  TGTGTATTCA  TGTGTATTCA  TTATTATTCA  TGTGTATTCA  TGTGTATTCA  TTATTATTCA  TTATTATTCA  TGTGTATTCA  TTATTATTCA  TTATTATTATTATA  TTATTATTCA  TTATTATTATATATA	ATAAGGATGA GCAATGTGAT ACAGTTTCG ATATCTGCTAT ACAGTTTCG ACACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTTAGGAG TITTAGTATG GATTAGAAGTTAGA GTTTAGGAG TITTAGTATG GATTAGAAGTGA AGTTAGGAG TGGTCACCT AGAGGACAGA TGGTTCACTC AGAGGACAGA ACATAGAAT CCTAATTACA CCTAATTACA CATTAGTAT TATACAGGAG AATATTCAG CATTAGCAT TCAGTTGAGA AGATTCAGT TATAGAGGAC AATATTCAG CATTAGCAT TATACAGGAG AATATTCAG AGAATCGCC AGTTAGCAT ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA AGAATCGCC AGTTAGCCCA AGTTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGAAA TGATTATCTC ATGTGCCCCTA ACCTTAGAAA TGATTATCTC TACACCACA TGATTATCTC TACACCACA TGATTATCTC TACACCACA TGATTATCTC TACACCACAC TACACCACACAC TACACCACACAC TACACCACACACA	ACCTGGCTA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA GATANATGT GATCAGGTAC ATTCGACAGA ATTCGACAGA ATTCGACAGA ATTCGACAGA GACTACAGA TANACTGGA TANACCTTGCA GATANAGAT TTGGACAGA TATCCCAGC GAGTTANTCA TANACTGAT GAGTAGGAGA TANACTGAT TANACTGAT ACTTGACAAA TANACTGAT ACTTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACGAC CATTGACCAT CACTCGACTC CACTGCACTC CACTGCACTC CACTGCACTC CACTGCACTC CACTTGACCAA TATAACAAA TACTCCCTTAG TACACACAA TACTCCCTAG TACACACACA TATTAACATA TATTATATAT	TEGAGCCTEG TTGAGCCAG TTGGCTTCTA TTAMACAGT TTGTCTAGGG CCACTAGGTA COTTCTCTT TTTTGTATGG CACTAGGTA AMATGCATT CGGAGGGGG AMAGAGGGG TTGATGTGAT AGCCCTTAG AMACTGAT ACCATGATA ACCATGATA ACCATGATA ACCATGATA TTAAGGATA ACAGGCTGT CCTAGCAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC CCACGGCCCAGTTC TCCAGGGCCA TTCACAGTCC CAAGGACTA TCCACGGCCAGTTC TCCAGGGCCA TTCACAGTCC CAAGGACTA TTCACTTGAG CCACCTGGGC CAAGGAGCAAG TCTTCACAGTC TCTTCAGGTC TTTCATGTT TCTTCAGGTC ATATTCATTA TTTCATCATT TTTCATCATT TTTCATCATT ATATTCATCATT ATATTCATTATATCATT ATATTCATTATTCATT ATATTCATTATTCATTATTCATTATTCATTATTCATTATT	GAMTANTGA AGAGANTTAG AAAGGAGTTT TCTGTAMAGT TACTAGGGGT TTACAGGGTT TTCGAGAGGTAT TCGAGCAGA AGTGAAATGA TCAGAGGAGTAT TCAGAGGAGTAT TCAGAGAGAACAGT TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1411 1511 1711 1911 2011 22111 22111 2311 2411 2511 2611 2711 2611 2711 3011 3111 3211 3311 3411 3713 4011 4011 4411 4511 4611 4611 4611 5111 5111 51
CCOTTGAGGA CGGTGAGGA CGGTGAGGA AGAGTATAAG ATTAGGGAAAA CGAAGTATAAG ACTATGACAA CGAAGTATAAG ATTATGACAA AGAAGTATAAG ATTATGACAA AGAAGTATAA AGAAGACACA TAGTACAGCA TATTAGGA AGAATATTAGGA AGAAGTATA TAGTACAGTA TAGTACAGTA TAGTACAGTA ACAGGAGTTA ACAGGAGTA ACAGGAGTATA ACAGGAGTA ACAGGAGTATA ACAGGAGTA ACAGGAGGA ACTTGAGAG ACTTGAGAG ACTTGAGAG ACTTGAGAG ACTTGAGAGT TAGGAGGACA ACTTGAGAGT TAGGAGGACCA ACTTGAGAGT TAGGAGGACCA ACTTGAGACT TAGGAGCACA ACTTGAGACT TAGGAGCACA ACTTGAGACT TAGGAGCAC ACTTATATATG	ACCIONALE ACCIONALE ACCIONAL A	AGTIMETT GGAATITITC GGTATGATT TATGCATCA AATGGGAAA GCAAGAAGG TACCACTGAT CACTATTGATC AATGGGAAA GGAATAGACTA AGGATATAGAC GTATGACTTA AAATAAGATT AAATAAGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGACGAC GTATGAGGAC TATCATTATT ACACCTGTGA AAATTATTCT ACACCTGTGA AAATTATTCT ACACCTGTGA AAATTATTCT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT TCCATGGTT ACACCTGTGAA ACACTGTAGAC ACTTGAAATG ACTTGAACAC ACTTGAACAC AACACTGTAG	TAGGATAGA TTAAGGATAGA AGATGTCCT TAAAGCAGCA AGATGTCCT TAAAGCAGCA AGATGTAGTG AGAGCTGGTC AGAGCTGGTC AGAGCTGGTC ACACTGTGG GCTCATTIAA ATTATCTGTG GCTCAGTATAA ATTATCTGT GCTCAGTAAA TTAAGATTATATAT TAGGTCATT AGAGATATT AGAGATATT AGCATCAT ACCCACAGAT TAGAGTACT ACCCACAGAT TAGAGTACT ACCCACAGAT TAGAGTACT ACCCACAGAT TAGAGTACT ACCCACAGAT TTAGACATCA ACTGAGGAAC GAACCTGGGAA ACAGCTAA CTGAACAGCT GGAACAGCT TGAACAGCT TGCACAGCTAA TCCATTCCTT TATAGGTTACT TATAGGT	TGGGGNAAGG AGATAATACT AGATATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGACATCA ATCAGATAATACT GGATGCCCAG GGAGACATAA ATCAGATACA TTTTGTGGGA CAGAGTCTGA AAACAGTAGT TTTTGTTGAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTGCCCTAC TGTATTATAAA CCTGATTTATAC TCAATTAAAA TGCAAATAAC TCAATTAACT TACAGGAGGA CACTATAAT GGGTGTAGAT GGGTTAAATT GGGTGTAGAT GGGTTAAATT CCAATGAGA CATGAAAAA CCAATGAAAA CCAATGAAC TACATGCGG TACATACCTA CAGTGGTTT TACATATACCTA TACATTACCTA TACATTACCTA TACATTACTATT	ATAGATAGA  AAGGAGAAG  GCAGCACTCT  TTATTATTGC  TTATTATTGC  TTATTGGGTT  TGAGAGGTTT  TGTACTGG  GCTTGCAAG  GCAGTAGACA  AGGGTAGACA  AGGGTAGACA  ATCCTGAATG  CTATTTGTGAG  ATCCTGAATG  CTATTTGTGAG  ATCCTGAATA  AGCCACTGT  AGGGTAGACA  AGGGCTAGGC  AGGGTTGCT  AGAGTAGTCA  AGGGTAGACA  AGGGTAGACA  TATGATATAC  CAATTGTCA  AGGGTAGGC  ACATTGTTAAA  CAATTGTTAAA  CAATTGTTAAA  CAATTATTCGG  CCCAGCACTT  AAAGTAATACACA  AAGGGAGGG  CCCAGCACTT  AAAGTAATACACA  CAATTGTTAAA  CAATTATTCGG  CCCAGCACTT  AAAGTAATACACA  TAAGGGAAGA  CAGGAAGATCCTTTATTAA  GCAGTGTGGC  CCTGTTATTA  ATTATTCTGG  GGTTGTATTCA  GTTGCACAAA  ATTATTCTGG  GGTTGTATTCA  ATTATTCTGG  GGTTGTATTCA  ATTATTCTGG  GGTTGTATTCA  ATTATTCTGA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TGTGTATTCA  TTATTATTCTA  TGTGTATTCA  TGTGTATTCA  TTATTATTCA  TGTGTATTCA  TGTGTATTCA  TTATTATTCA  TTATTATTCA  TGTGTATTCA  TTATTATTCA  TTATTATTATTATA  TTATTATTCA  TTATTATTATATATA	ATAAGGATGA GCAATGTGAT ACAGTTTCG ATATCTGCTAT ACAGTTTCG ACACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTTAGGAG TITTAGTATG GATTAGAAGTTAGA GTTTAGGAG TITTAGTATG GATTAGAAGTGA AGTTAGGAG TGGTCACCT AGAGGACAGA TGGTTCACTC AGAGGACAGA ACATAGAAT CCTAATTACA CCTAATTACA CATTAGTAT TATACAGGAG AATATTCAG CATTAGCAT TCAGTTGAGA AGATTCAGT TATAGAGGAC AATATTCAG CATTAGCAT TATACAGGAG AATATTCAG AGAATCGCC AGTTAGCAT ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA AGAATCGCC AGTTAGCCCA AGTTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGTA ACCTTAGAAA TGATTATCTC ATGTGCCCCTA ACCTTAGAAA TGATTATCTC TACACCACA TGATTATCTC TACACCACA TGATTATCTC TACACCACA TGATTATCTC TACACCACAC TACACCACACAC TACACCACACAC TACACCACACACA	ACCTGGCTA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA TANACTCTCA GATANATGT GATCAGGTAC ATTCGACAGA ATTCGACAGA ATTCGACAGA ATTCGACAGA GACTACAGA TANACTGGA TANACCTTGCA GATANAGAT TTGGACAGA TATCCCAGC GAGTTANTCA TANACTGAT GAGTAGGAGA TANACTGAT TANACTGAT ACTTGACAAA TANACTGAT ACTTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACAAA TATTGACGAC CATTGACCAT CACTCGACTC CACTGCACTC CACTGCACTC CACTGCACTC CACTGCACTC CACTTGACCAA TATAACAAA TACTCCCTTAG TACACACAA TACTCCCTAG TACACACACA TATTAACATA TATTATATAT	TEGAGCCTEG TTGAGCCAG TTGGCTTCTA TTAMACAGT TTGTCTAGGG CCACTAGGTA COTTCTCTT TTTTGTATGG CACTAGGTA AMATGCATT CGGAGGGGG AMAGAGGGG TTGATGTGAT AGCCCTTAG AMACTGAT ACCATGATA ACCATGATA ACCATGATA ACCATGATA TTAAGGATA ACAGGCTGT CCTAGCAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC TTAGAATTAAGGTTC CCACGGCCCAGTTC TCCAGGGCCA TTCACAGTCC CAAGGACTA TCCACGGCCAGTTC TCCAGGGCCA TTCACAGTCC CAAGGACTA TTCACTTGAG CCACCTGGGC CAAGGAGCAAG TCTTCACAGTC TCTTCAGGTC TTTCATGTT TCTTCAGGTC ATATTCATTA TTTCATCATT TTTCATCATT TTTCATCATT ATATTCATCATT ATATTCATTATATCATT ATATTCATTATTCATT ATATTCATTATTCATTATTCATTATTCATTATTCATTATT	GAMTANTGA AGAGANTTAG AAAGGAGTTT TCTGTAMAGT TACTAGGGGT TTACAGGGTT TTCGAGAGGTAT TCGAGCAGA AGTGAAATGA TCAGAGGAGTAT TCAGAGGAGTAT TCAGAGAGAACAGT TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1411 1511 1711 1811 2011 2011 2211 2211 2211 2311 2311 23
CCOTTGAGGA CGGTGAGTGA ATAGGGATAA ATAGGGATAA ATAGGGAAAG CGAATATAAG ACTATGACAA ACTAGGAGAA CGAAGTATAAG ACTATGACAA ATATAAGGAAAG ATATTAAGG ACTATGACTA TGTATCTCC TGTACATCT TGTACTTGTA TGAAAAGGA ACAGGAGTTA ACAGGAGTTA ACAGGAGTA ACACAAGA ACACAGAG ACTTGCTCA CGACAAACA ATAGGGGTAT TGCTCAGCA ACACAAACA ATAGGAGTA ACACAAACA ATAGGAGTA ACACAAACA ATAGGAGTA ATAGGACTA ATAGGACTA ATAGGACTA ATAGGACTA ATAGGACTA ATAGGACTA ATAGGACTA ATAGGACCA ACTACACATCT ATAGGACCA ATAGACCA ATAGGACCA ATAGGACC	ACGOGCEAGE TITGCIGAGA TITGCIGAGA TITGCIGAGA ACCAGACTCC GAACACATTA GGCATITTAT GGGAGACAGG GCACTCCIGA AGTAGAGAGA CTAAGAGAAC CTAAGAGAAC TAGGTGAGAACAT TAGGTGAGAACAT TAGGTGAGAACAT TAGGTGAGAACAT TAGAGAACAT TAGGTGAGTA AAAGCTGAG GATCAAGAGAC CTCATTTATAC AAGTACAAAGCTGAG CATAAACTCAT TAAAGAGCT CATCTTTTATC AAGTACAAAG CCTCCTTTTATC AAGTACAAAG CCTCCTTTTATC AAGTACAAAG CCCACCAATT GAGAGGTGAG GATAAAAAAAA CTCGTTTTATC AAGTACAAAG CCCTCTAAGAA CTCCTTTTTATC CCTTTTTATC AAGTACAAAG CCCTTCAAGAA CTCCTTTTTATC CCTTTTTATC AAGTACAAAA CTCCTTTTTATC CCTTTTTTATC AAGTACAAAA CTCCTTTTTTTTATC AAGTACAACA CCCTTGAAGA CATGTGTGTATA TATATGTATA	AGGITMETT GGAATTITTC GGTTTGCATT PATGCATCAA CTCTTTGATC AATGGGAAA GCAAAGAAGG TACCACTGAT GACTATGCA GTATGACTTA AATAAGATG TACATTATACA CTATGACTTA AAATAAGATG TACATATACA CTGGTTGTAT AAATAAGATG AGGTGAGACT CAGGTGACT AAAGAGACT AAAGAGACT AAAGAGACT GAGGTCAGA TCCATGAGAC AGCCTGGAA AGCATCCAT AAGGACT AGAGTCCAC AGCCTGGAA AGCATGAGT AGAGTCCAC AGCCTGGAA AGCATGAGT AGATGCAC AGCCTGGAA AGCATGAATG AGCATGAATG AGTTGAAGAG ACTCATAAA AAGGAACAC AAGGAACAC AAGGAACAC AAGGAACAC AACCTGTAG AAGCAACAC AAGGAACAC AAGGAACAC AAGGAACAC AACCTGTAG AACCTTAAATGG AACCTGTAA	TAGGATACAC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAGGATAGA ATCATCGTG AGAGCTGGTC GGCTGATTTA ATTATCTGCG ACCATCTGGG GTCTGATTTA ATTATCTGCG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ACCATCTGGG ATTAGCTATTA ATTATCTCTGT GGCAGAAACCC ATAGGATGAT TTAGGTCATT AGGAGTATAT AGCCACAGT ACCACAGTATA ACCAGGAAACTA CCCTAAACAA TTAGGTTATA GGAAACTAC GAACCTGGGA ACCACGGCATAACAA TCCAATTAGA GTGAAACCCC GAACCTGGGA ACCAGGCATCAA GTGAACCCC GGAACTACT TATAGCTTAT TATAGCTTAT TATAGCTTAT TATAGCTTAT TATAGCTAT TATAGCTAT TATAGCTACA ACTAGGAAC CCCACAGCTAAACAA TCCAATTAGCTAA CTCAAGCAAC TCCAATTATAGCTAA TCAATTAGCTAA TCAATTAGCTAA TCAATTATAGCTAA CACATTATATGCTAA	TGGGGNANGG AGATANATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGACTCA ATCAGATATA CAAAGACTCA ATCAGATATA TGCTCTTGGGGA TATGGAGGCCAG GGAGACATAA TATGGAGGG TTTTGTTGGGA AAACAGTAGT TTTTGTTGTGA AAACAGTAGT TATTATAAA CCTCATTGAT CAAACAGAAG TGATTCAAAC TGATTGAAAT TATTATACT GGCTTAAAT TATGGGGG GGGCTAAAT GGCTTAAAT TATGGGGGG TACAATACTA AACAGTAGG TACAATACTA ACAATGGAGGT TACAATACTA ACAATGCGG TACAATACCTA ACAATGCGG TACAATACTA CAATGCGTTT AAAATGCGTTT AAAATGCGTTT TACAATACTA TATATATA	ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATTOC TTATTGAGTT TGAGAGGTTT TTATCACTG GCTTTGCAAG GCATTGCAAG GCAGTAGAC AAGGCTAGAC CAAATAGCTA AGGCTAGAG TATCATATATAC TACAGCACGT AGGGCCATGG AGGGTCAGGG AGAGTTCCT ACAATAGTCA ACATAGTCA ACATAGTCA ACATAGTCA ACATAGTCA ACATAGTCA ACATAGTCA ACATAGTCA ACATAGTCA ACATAGTCA CAATAGTCA ACATAGTCA ACATAGTCA ACATAGTCA CAATAGTCA ACATAGTCA ACATAGTCA CAATAGTCA ACATAGTCA ACACTATTA ACACTATTA ACATAGTCA CCACTAGTCA CCACTAGTCA TGCAGCAGCA ACAGGAAGAT CCCGGTTGTATA ATATATCTG GGTTGTATTA ATATATCTG GGTTGTATTA ATATATCTG GGTTGTATTA ATATATTCTG GGTTGTATTA ATATATATATA TGTACACACA	ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACACTACGG CAATATAGG GGACTTAAGAA CGGACTAGGAG CGCTAATGAA CTGAACAGTA GAGTCTTGGG TTTTAGTATG GATCATGGG GTGTTGGGCAT CCGGACTAGA TGGTTCACTC AGAGGAACAGTA AACATGAAG CATAAGGAG TTTCATTATA AACATGAAG CATATTCATC CATAGCATT TATATCAGAG GAGACCACT TATATCAGAG AGTTAGCTG AGATGGCAT ATTTGGTG GACAGCATT TATATCAGAG ATTTAGCTG AGAACACCA AGTTTAGTT ATTTATACA AACATTAAAG ATTTTATACA CATAGCATT TATATCAGAA TGATTAGTT ATTTTAACA AACATTAAAG ATTTTTATAC AAGAATCTGT ATTTTGGTG AACATCACA ATTTTTAACA AACATTAAAG ATTTTTAACA AACATTAAGAA TGATTATCTT TATATGAAA TGATTATCTT TATATGAAA TGATTATATA	ACCTGGCTT TCTGCAGCA TANACTCTCA NATTTTGANATGT GATGAGTAG GATANATGT GATGAGTAG ATTCGAGAG ATTCGAGAGA ATTCGACAA ATTCGACAA TAGATCGTG GATANATGT TTACCTGCA GATTATGACAA GACTTTGTGT TTACCTGCA GATTATGAA TAGATAGGAT CTGGAGAGTG GAGTTAGGAA TAGATAGGAT TAGATAGGAT TATATGACAA TATATGACAA AAAAAAAAAA	TGGAGCCTGG TTGGAGCCAG TTGGCTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA AGGTTCTCT TTTTGTATGG AAAATGCAT TGGATGGAGA AAAGAGGGG TTGATGTATA AGCATGGATA AGCATGGATA AGCATGGATA AGCATGGATA AGCATGGATA ATTCAAGTTAC TTAAGGATGA TTAAGGATGA TTAAGGATGA AAAAGAGAT AACACAGATCT CAAGGCCTGT TCTAAGGTTC GTTATTTAAG TTAAGGTTC GAAGGAAGTA AGCATGTTCA TTAAGGATGA AGCATGTTCA TTAAGGATC TTAAAAGAGAT TCTAAGGTC GAAGGAAATT ATTCAAGTC GAAGGAAATT ATTCAAGTC CAACAGATCT TCCAGGGCCA TTCAAGGTC TCCAGGGCCA TTCAAGTC CAACAGTC CAACATTAAG TTTCATCATT TCTTGAGATG CACACATATT CATCATGGAT ATTTCATCATT CATCATGGAT ATTTCATCATT CATCATCATT CATCATCATT CATCATCATT CACACACA	GAMTANTGA AGABATTAG ALAGGAGTTT TCTGTAMAGT TACTAGGGGT TTALAGGGTT TTTCATATGG ATGGAGGTAT TCTGAAATG TATGGAGGTAT TATGGAGGATG TATGGAGGATG TATGGAGGAGGAT AGAGCACAGT TCAGGAGGAGGAT ATGAGAGAGGATAAA ATATATAAGG TGAGAGGATAA ACAGCATGA ACACACAC AGAAATATTGC GAGAAATATGG GAGAAATATGG GAGAAATATGG TCAGGAGGAT AATCCAGCATA AGAGCTTGTG TCAGAACATGA AGAAATATGG TCAGAACATGA AGAAATATGG TCAGAACATGA AATCCAGGATGT AATCCAGGATGT TCAGAAATATGG TCAGAACATGA AATCCAGGATGT TCAGAAATATTGC TCAGAACATGAAA TCCAGACATT TCAGATATATATT TCGATATATATT TCGATATATATT TCGATATATATT TCGATATATATAT TCATATATATATATATATATATATATATA	1411 1511 1711 1911 2011 2211 2311 2411 2511 2611 2711 2911 3011 3111 3111 3411 3511 3411 3511 4011 4111 4411 4511 4511 5511 5511 5

### Figure 8B

AGATGTAGCT	TANANANA	ATATCCTGGA	ATTCTAGAGA C	ATCCTTAAA	TCACTGCAAT	TCCTATAACA	CTTGCCAACC	AAAGGTGCTG	TTGATCTGAA	581i
· concretelist	CITCAAAGTC	ATTTCCCCAA	CCATATICIS A	AAACAGCCC A	AGCCAGGGTG	ATGGATCACT '	TTCCALLGIT :	こうかいしょう しょうしょう こうしゅう しょうしゅう しゅうしゅう しゅう	COR & STREET, CA &	
										6111
GAAATTGGCT	TICAGATTAT	TIGGATIAAA	TAATTCCTAA A	TCTTAAGAG	ATY:T2222TT	TTCATCATCT '			CCCTAAAGA	6211
							-1 -1			6377
CTTTTACATT	VAI PD	e Leu Ard K	IR GIU ARD AI	a lon Ive	TIA TAN BAS	. Rea Dea Iv	~ l - ~ This l	6 61	Lys Leu	
1	•		AT GAA AAC GO		14-		0 V00 IVI V	AT ICK OUT	AAA TTG	6393
OTA OTA SU	e Agt GTU A	TA WRU TER	GIO NEG GIO C	VR Met. Cin	CIN LVe C	JO SAT DRA G	in Gin Ala	Arm Clu Wal	Db - Cl	
	38	: ":	and han the		ann nno 1	Francisco	nn unn uch	• • • • • •		6474
Asn Thr Gl	u Arg Thr		CACATAATAC (	CTTC CARC				~		
									·	6570
AAAGCATCCA	. TATATATTTA 9	TGTATGTTAA	ATGTTATAAA I	GATAGGAAA	TCAATACCAA	AACACTTTAG	ATATTACCGT	TAATTTGTCT	TCTTTTATTC	6670
n th	r Glu Phe 1	rp Lys Gln	Tyr Val	1.56 B. 1885	in the same of the				Section .	٠,٠
TTTATAG	T GAA TTT I	CC ANG CAG	TYP Val	PAAGCAATTC	ATTITATECT	CTAGCTAATA	TATGAAACAT	ATGAGAATTA	TGTGGGTTTT	6763
TTCTCTGCAT	. AAATAGATA	TATATTAAAC	TTTGTCAAAA (	GACTCAGAA	AGATCAGTCC	AACCCTCTAA	CCCATATTGG	ATGGTGATAT	ACTACAGGGT	6863
TATGCCAGT	i TGGGAACTAI	CGCTGGTAAA	TAAGTTTAAT (	CTCCCTAGG	CCTTCACAAA	GAACATTGTT	CCACCCCAGG	ACCOMOCARO	CARCARACTO	6963
TATCATGGCC	: CAGAAGCCC	TCCATGATTG	TTTGCATTCC TCCTTCCCCA	<b>・アイフンアンファ</b>	<b>でょっととうかんかか</b>	CCCTCCTCTC	CT & CTTCTCTCT	CCCCCCCC		7063
CCACCCTGG	CTTCCTGTC	CTTCTTGCAC	ACTCTAGGAA TTTGCCCAAA	CCTCCCACT	TTGGAGGCTT	TATCTGGCTG	TTTCTCTTAT	TIGGCTGTIC	CCAACTTCCT	7163 7263
GCCTTACCAC	TCTAGACAC	TGTACAGAAC	TCCACTCTAC	TTTTAACAG	TTAATGAGGC	CTACCTTCAC	CATCTATTAA	TACTTCAACC	TGCCCCAGTA	7363
CATTTATTT	CTCCTACTC	: ACTAAAATGC	AAGTTTCATC *	PTGGCAGGGA	TATTC ARTIC	Andread Calculut V th	TO ATATA TOTAL	CTLCCLCCTL	CALCACRAGO	7463 7563
TGGAAAAGAG	; GTACTCAGT/	<b>\ AATATTTATC</b>	AAATGAATTG	<b>NCCAXAAGAA</b>	GGAAAACTCA	AAACTTTAAT	GACAACTAAC	TTTTABACCTA	CARTIROPES .	7669
CCAACTTCC	N TCAATIGGA	A ATCCAATICE	AGTTTCAAAG ATTTTCTACA	GTTTATGTTC	TGGAGACACT	ACTGGACACT	CTTTTACTCT	CATARCTCAT	AACTCCTCCA	7763 7863
CTTTTGTTT	TAAATCATG	A GAGAAAAAGA	GTTGACTCTC	TTATATTCTT	<b>マサムヤイサルイイツ</b>	ALK CALLO ALL	TTACABACCA	ATACTACCAT	ACCACCOMOCH	7963
CCATCCCCA:	r TTTCCCCAC	CAAAAAATTCC	AGGTCTTTGC TGACTATTAA	ACTCCTACAA	TCCCTTCATT	GCTCACTCCC	CACCCCCAGG		A A A COROCCCC	8063
CCTTGCCTT	r TGGGTCACA	I AGGTACACTG	TITIGCTATAC	CACAGGTATA	COTATORCA	AAACATCCAC	CCTATTATTC	TOTAL COLUMN	COTTO COTO A B	8163 8263
CCAAAAAA	N AAACAAAAC	N AGAACAAAAA	AGAAACAAAC TCACTGCTGT	TCCCTGCCTC	TTTTCACTTG	CACTCAACGT	TCCTAACCAC	TACALASTY	CCCTATCTT	8363
ACATTACTA	G CATCICICA	A GACTITCCAG	TTTACAAAAG	GCCTATCACA	TTTAACCCTC	<b>ACACCATYC</b>	TCTCACCAAA	CCATTATAA	CTCC) TOTAL	8463 8563
CAGGAGAGT	N AACTGAAGC	I TAGGGAAGTI	GGGCCTTCAA	CCAAAGGTCT	CCCAGTTGGG	GACTCATGAA	GCCCAGAAGA	GAAGCCAAAT	TOTOTOTO	8663
TGGTAAGAT	C TIGICICIN	G TTGTATTTGA	CCCCAACTGT	CTATGGCTTT	GCCTGAACCC	AAAGTACACA	CAGCCTAGAA	ACCARAGGAG	AACCAAATCT	8763 8863
GGGATAAAA	I GACACTCAT	I TTAACGACAT	GTCTCAGCAA	ATGAGTTCCT	GTGTAGCTGG	CTGAAAGCCC	AGACCCTTTC	ACTABABCAT	CCTCAATAAT	8963
CATAGAATT	G GAACAAATT	A GAGTATCTGT	AAATGTAGCT GCAAAAGCAT	CATITITAGA ATCAGATCTA	GCAGCAGAGG	CONTONATAGE GGACAAGGTC	TAACAAACCA	GAGATAACCG	ATTTTGTTTT	9063· 9163
CGGACTACT	T ATGATAAAG	G GATATTAGTO	TCTTAGTCAA	CGGAACCTGG	ATACACGCTT	CTGACAGAGA	AGAGGGAGAA	TACCCACCAA	TOTACACACC	9263
AGATGTCAA	G GAGATTTGC T TGTTTCTCA	T TTAAAATACG T ATATTGAGTG	ACTGATAATT TTACAGATCA	AGAAATTTCT AGCTCCCATT	CAGTTTCCCC	CTTTTCCCTC	ATTCTTTGAT	TCTTATTGTT	ATCTTTATCT	9363
TEGITETAT	T CATCCITCE	L TUUCAAAGC	r CCTTTAGAAG	TOTOGRAPHAL	CCCACACCA	TARCARACCA	CACTTABACA	dated Construction	The state of the s	9463 9563
TICICCTIT	C ACCTATTCC	T TOUTCOTGET	TTCTTACCAT A GCACACTGTA	CAGTGTCTTC	AAAGGCTTTC	AAGTACACGG	TAAATGCAGA	AACTTCAAGA	AAGGCAGAAT	9663
TGAAACCGA	A TTTGTAAAA	C ATAGACTATY	CITTAAGTAGT	AACAGATGCT	TCTGACATGT	TITCTATIGE	CTTGAACCAT	TACTGCATAT	GATACATCAA	9763 <sup>.</sup> 9863
AGTTAAGTG	<b>A CANTACAAG</b>	A AAGCAGATT(	CATTTGCTCCC	TGCCTAGGCC	GTCAGTTCCT	* AAAGTGGAAA	CGCCATATAT	TATCTACCTC	AGTTTGCTCT	9963
ATTGAAAAT	C CTTCTCACC	C TGTGCTGAT	A TAGAGATAAT	CTATACAAAA	ACGTCCTTCT	CCCTCTTCCA	TTGGATTGCA	TANACTATET	ATAAGGAAGG ACATGCCTTC	10063
CTCRGGGGG	A CTITUTAG	& ACAGTGTCA	G CCTAAGGATC	TITGITTGGG	TGGCTTTTAG	; AAACTCAGGA	AGACAGGAGC	ATCATATGCC	TATAGGCAGC	10263
100C11CCA	o olcaciaci		C CCTAAAATCA 47					•		10363
C11-C11-C	-	A	sp Gly Asp Gl	n Cys Glu	Ser Asn Pro	Cys Leu As	n Gly Gly S	er Cys Lys	Asp Asp Ile	
CANTICANI	T TCTTAACCT	A TCTCAAAG	AT GGA GAT CA	G TGT GAG	TCC AAT CC	N TOT TTA AA	T GGC GGC A	GT TGC AAG	GAT GAC ATT	10450
Asn Ser 1	yr Glu Cys	Trp Cys Pro	Phe Gly Phe	Glu Gly Ly	s Asn Cys (	lu Leu				
AAT TOC 1	AT GAA TGT	TUG TGT CCC	TTT GGÅ TTT	GAA GGA AA	G AAC TGT (	AA TTA GGT	AAGTAACTAT	TITITGAATA	CTCATGGTTC	10537
AAAGTTTCC	C TCTGAAACA	A GTTGAAACT	G GAAAATGCAA	TATTGGTGTA	TCATAATTT	TCTTAAAAAC	ATACCTTTGA	TGCTTATAAA	CATTTCATTT	
TATTAATCI	T TANTACAAC	C GTATGTGGT	A GAAGCTACAT T AGTACTATCA	TAXAATCAAT	TCTATGCAG	I GGTAACTAAT A TGAGAAAACC	GCAACTCCAA	CGGCCAAAAA	CGACTCACCT	10737
ATAAATGGT	T TAGAÇAGGA	IC TTAXACTTC	A GTGTGACCAA	AACCCATGCT	TCTAACTAC:	r atattcaaaa	CTCAGAGAAA	<b>ACTGAACCCA</b>	GAAAATTGAA	10937
ACAAACCAC	TA AATTGCTAI	C AACATAGGT A AACTTGTTG	G AAAGTCAATT A CAACATTCAA	AAGTACAGAA	CTGGAGTATCA	G ACTGGCCAAT	TATCCCATAT	AATGGGAATT	CTCCACATGT	11037
ATCCTTGTC	T ACCITITC	T CTCAAAGCC	T AGATTATITC	TTTTTCCGAC	GTTTTCAGT	N ATTGGAGCAG	TAAACCCCAG	TGTCCCTTAC	CTACTTGTTT	11237
ATTACCTCC	IA GATGCAATA BA AGGAAGAAT	NT TACTGGTAC	T GTGATTGAGA A GGCTACATGA	ANCGCACACA	GTGCTAATG	A GGAATTCACT	TTCTACTCTG	ACACTCTGGA	AGAATAGAGA	11337
ACAGATGA	<b>LA AAGGCATG</b>	la atgaaatga	A ATGTAGCAGC	TACACTCGTC	: CTATTGAGA	A AGGAAAAAAG	TCACCTGTAA	TGTTGTTCAC	AAATCCTTTC	11537
AGTACTAA	N ANTICATIO	IA CCATCTTCC	T TTAGTCTCGA	<b>あみみすすすぐすする</b>	GRACCTARA	R BRECCERER	CTCACACGGC	LAAGACATT	CHARLESIAG	11637
TGTGTTAA	T GATTCAATA	T CTCTGAAGT	G CTACTTTCAT	CTGAAAGGT	ATAATTTGA	A ATTCAGATTI	ACCTGGATAA	ATTTGATCT	GCTATTATGG	11837
AAACCTCT	NG AAATCCTT(	G AGTAGTTAC	T CATTATCAGC	TTAAATAATA	<b>TAGCCGGTG</b>	G AGCTGAGGGA	ATGAGTAACT	CAATTAGTCT	CAGTTACAAC	11937
GAATGTAA	AT ATANGCAC	M ACTATAATT	G AAAACATAAA A ACAGAGTTTG	CTACGTGTGT	GGCTGTGTAAA	A ANTANAGATG C CACCCAGCAA	AAATGCTAAG	AGGAAGGGAI	TAGCCCTGAG	12037
GATACTCT	CA TOTTCCCAC	CA TITTGGTTT	G GTCAAGGCTG	TGCAGTTGT	CTGCAGGCC	A CCACCACTCC	TGGCCTCTAC	AGTATATTG	TCTGACCCAC	12237
CAATCTGA:	IV ANGGTTTA( MA ANGCATAL)	TTTATAAA AU ACAAAAAAAAA	T CAGCCCAGTT	AGCTCACAA!	CAAAATGAG	A TTGCAACAA	ATTGCTCTTT	ATCTCAGAC	ACAGAGGAAA	12337
GGCCCCAA	GT CTTTTGTC:	IT ATAAGGTCI	T GAAAAAAA T	AAGGAGATT	TCATCAATA	A GAGTTTTTT	TTATCTTTT	CCCTTGTTC	TCAGGCCCTT	12537
CACTGCGA	GA GAGAGGTG:	IN ANCETTCHE	ig gcatgcattc	TAGTTAAAGI	TTAATTAATT	G CCTATTGGG1	CCCTTTCGTT	RGAATAAAGI	CCTCTGTATG	12637
TTCCCGCA	GT AAAGTAGA	AT GGAAAGAAA	IA TCTCTCAGAT VA CAAAAATCAC	AAGCCTATAL	L ACACCTTCT	T CAATTITCCC	AGCATGTCAC	* AGACACTAC	CTCTTATTA	12837
CTACGTAT	TT CTGAGGAG	TA AAAAAAGGA	LA ATATGTTGAG	TITAGCTGA	A GCACAGCAT	A TITTGTGGT	<b>AACTTGTTA</b>	ATABABCAT	TTTTCTCCAA	1 2937
AAGTGTTT	GT GTTAACAT	التكفافا كالمكافاتا كالماليا	A GGAAATATAA NG AACAGATGCT	TITUALIATUT	r CTTCAACAG	A TATTCTACC	: ACTGAGACCG	COTTCGGGM	CACEGAGAGC	1313
CCATATCC	AC CACAGTAC	CT GACACATA	A TGCTCAGTAA	TTGATAAAT	AGTCCCATT	C TAACTGTTC	TTAGCCCTG	TCTATGGAA	TCTCCCTGA	13237
CAAGOCAA	AG ATCAACAG	CA GCAGCAAC	AA TGCTCAGTAA AA TCTTCAGCCT AT ACTGAGCCCT	ANAGGGCAN	T GACAAAAGA	CAGAATGATIA	GAAGCAATAT A CAGAGGTCT	TTCCCACCT	CTGCGCAAAA I AGCCAATGAC	1343;
ACAGAATC	AC AATTGAGA	AA ACACAGAG	TATTCATTCC	CATTOTOCA	CCCTGGAC	A AACCAAGCT	CACCTITCG	T AACTTATCA	C AATCTCATAT	1353;

## Figure 8C

TEACOGRACA CTITCTACAG GTAATGITTG ATTTGGCTGA ACACTITAGC ATTGCTTCGT AGCAACAAAA TGATAGCTAG TAACAGAAAA AGATCCAGGG	1363
ANATIACCAC TGTTAGTGAG GAGAAAGGCC TTTTAATTAA TTAATTAATT AATTAATAG ACCAAGTGCC ATCTTTTTGG ATCATGCCCT TAGTGGATTA	1373
TIGGTAGGAA AGGTTAAAGC TCAAGGTGGT TCCTTTGTCC CCCTGGCAAC AGTTGATTTG CCTCCCTTAT CTCCTGAAGT ACCGTAAGGA CTAAGAGCCA	1383
ATTATTACAT TOGGTATGG TAGGATATGT AMATAGAGT TIMANGTTT AGATTCATCA CTCAAAAATT CATATTCTCC AMACCATAC AGTCACTCTC TIMCCCTGTG TTCCCCCAGA AMAMAGTCA CAAGCTTATT ATTAACATGT GCAATCCAGG GGCAAGAGAA AGGAACTGAA GATGAGGCAG AMAGGAAAG	1393
ALAGCIANTA AGAGGATGAG TIATCALACT ACTIGITTCT TAACAGCALC TGATTGCTTA ACTICCTGGG ACTGTCTCCA ATAAGTCALA TIGGCCTCAG	1403
GITAGTCCAC CTGAGTOGGA AGAAGGGGTG AAAGAATTTG TCTGTCAGTA TCTGTCTCTC ATTGGTTAGA AGTTCGACTT ATGGGGAATT AACTCCTCA	1413
CATTILLIAG ILGGATAGAT TUGGTALLAG AGGCATATUS CATCTATUST FAUCATUAN ACCURACITY FALCONARA CAPARAMAN ACCARACITA	1423.
	1433.
	1443
CTUTICILAR AURUSTACAC RAGGGGRARG GTGTCTTACA TTYCTTATCT TYCTYCTTC CTYCTACAAAC ACAAAAACACAACAACAACAACAACAACAACAA	1453
TITANTIVA GUAGGIUTAN LEITANGARIT CUTGANATAT CETACTOTET COTATOCCETA TOCCATACCA CARACARIA CARACARIA CARACARIA CA	1463°
	1483
CACITAAAAA ITTAATTUTA TACTUTTAAC GAAAGTUATA ECTAAAATAA AATTACACTE EGACECCAAA ATCAACCAT COCAAAAA COCAA	1493
	1503:
AGGGGATT GCATAGGAGA GTANAGAAA TOTGGGCCAC TGGAATGCTT AGCACTANTG ACATATTGGT CITTGGTCTT CAGTACCTT ACAGGACCCT ATTCATTCT CITTATGTTT CATTACCTT ACAGGACCCT ATTCATTCT CITTATGTTTG ATATGTAAC ACCTCAGCCA GCTTCAAGTT GCTTTTTGGC CCTAATGGAC TTCCTAGCAC TATAATTTCT TITTTTTTTTAA	1513:
ATTENTED TENESTED AND TAKE ACCIDENCE SCHEMICH SCHEMICH SCHARGE TECTAGENE TRANSPORT TENTAL TENTAL ACCIDENCE TRANSPORT TENTAL ACCIDENCE ACAGGGGTT STIGTACATA TENTACATE ACGCAGATAT	1523;
	1533:
ATABUTUT ANGUTTAGE TUUGGTTAG AAGTGAGAAG ETGGAGTATT TGATTTTTGT TCCTAGGGTA GTTTCTAAG GARGAGAAG AGGAGGAAG.	1543;
TICATATICE CACAAAAGAC ATAATCTCCT TCTTTTCTAT CCCTGCATAA TATTCCATGC TATATATCAA CCACAAAAGAC ATAATCTCCT TCTTTTCTAT CCCTGCATAA TATTCCATGC TATATATCAA CCACAAAAAAAAAA	1553;
TGGGGATITA GGITGATICE AIGICIGCIA TICTAACACT GIAATITCIA AAGACTICCA GATTCIACTT TIATAGCTAA CORCERAAC ACTUM	1563;
TGGAAGUUAA GUAATITUTA GAATAACTAA GCAATAGAAA TTACACTTCA ATGCAGAAAG GCAGTATCTA CATGAGATTA TGAAATTCCC CORCONOMIA	15737 15837
CTGITCACTU AAAAAAATAA GTAAAACTGT AACTTTCAGA AAAAATCATT GTACATATAG AAAACTGA GCATCTAAAC AATTAAAAAAAAAA	15937
GAANGATTAL TOWATAGA GTCAACATAC AAATATCAAT YGTATGTCTA TITACCAGCI ACGITTCAAL IITGITTTTT IRIIRIACAG RIIIIIAACATA	16037
ATAGGCANTG TITATCATTA AAGGATACAA TATAGTAAAT ATATCAAATG TITACTAATG GATTCAATGC AATACCAAAG TCCCAGCAGG CTTTTTTGGT	16237
COTOCGACGT COCCCAGGAT TCATAAGCTA ATTATAAAAT GCATATGGAA ATGCAAAGAG CCAAGGATAG CCAAGACAGT TTTGAGGAAG AATAAACTTG	16337
IACTACTIAC ACTACCAGAT GTCAAGACTT ATTATCGAGT TACATTTATT ANGACAGTGT GGTACTACAC CAAGGATGAT CAAATAGATC ACTGAACAC ACTAGAGGC TCAGAAGCAC ACCTGTACAA ATTATCAGAT ATTATCAGAT TACATTATA ATAGAGGTGC CAGTGCAGTA GAGAAGGAAA TTATTGGTGT TTTCAATAAA	16437
MOTGATAGG TCAATTAGAT ATTCATATGG CATGAAGTAT GAAACAATAA CAATTATAT TCATAACTTG CAGAAAGCAA AAATTCTTA AAATACAAA	16537
METGATURUS ATRANGUARA AGATTGATAA ACTGGACTAT ATTALALETA AGGACTECTEG TYCLGCALAL GACACTATECTO COLONGALALA CACALOGA	
MAGTUAUAL ARUATATUTU LAATACAGAT ACCTAATAAC TG11/CC1T 1/2/2/2/2/2/2/2/2/2/2/2/2/2/2/2/2/2/2/2	
WEAGTATUT ACTAGATUTG AACATGTGAT CCAGTAATTA CACTCATAAT TATAAGCCAG TAAAAAAGCCA TCTTTATCTC ACCAAAACAG AGAGAACAA C	7/033
ATUTICATTA CACTATIATA CATAAGAGCC AAAAACTGGA AACAAACCAA ATATCCATTA ACAGTAGAAT GAATAAATAA AACCTALE ICTAATAAAC	77007
IGGANTACTA CACAGCAATG TAAATGAACT ACTGCTGTAC AAAACAACAT GGTTTAATCT CACAGACAAA ATGTTAAATG AAAGACAG ACGAGTACAG	17117
ITCCGARCT TUTGTITATA ATTUARGARC TGGCRAGRAC TGTTTACTCT CTTRGARGYC CRGCTRATGG TRACCTRTR RARGERRAL CCCCCCRACC	17777
ITTGGGAGGG GGCATCITCI GGGGTCITGA TARTCTGCTA TCTATTGCTC ACTTTACTCT TTALLCAGGC TCATTTACTT TCTGALALCT TACAGALALCT	17554
HIGHTOTAT TITTEGATA TATGETATAC ATTACTAAT AGGSTTTTA AACCTAGT TCATAATTTA GTGAAAGTAG AATATCCAA CATTAGTTT	17437
TANCCANTE ANTINTAGES CTACCATENT TITTATGCAT TATTGAGANG TITATTTTAC CTTTCTTTCC ACTCTTATTT CANGGCTCCA ANATTECTCT	17537
CCCANCGEN TATEGGGGG ANCATGANTG CCCCCANTGE ATATETGAGC CATACATGAG TCAGENGETC CATGENCETT TENGRANGG ATGETALATG	17637
Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe Cys Lys Asn Ser Ala	
	17724
MP ASS LYB 'Val Val Cys Ser Cys The Glu Gly Tyr Arg Leu Als Glu Ass Gls Lys Ser Cys Glu Pro Als GAT ANC ANG GTG GTT TGC TGC TGT ACT GAG GGA TAT CGA CTT GCA GAA ANC CAG NAG TGC TGT GAA CCA GCA GCA GTCATAATCT	
GAT AAC AAG GTG GTT TGC TCC TGT ACT GAG GGA TAT CGA CTT GCA GAA AAC CAG AAG TCC TGT GAA CCA GCA G. GTCATAATCT	17807
•	
CANAGATE TITTAAGAA AATCIGTATC TGAAACITCA GCATTITAAC AAACCTACAT AATTITAATT CCTACTIGAA TCTGCTTCCT TITGAAATCA	17907
INCHARANTAT CAGTAGCTTG AATTAGACCA ATTAATTTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATT ACAACCACA TTAATTTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATT ACAACCACA TTAATTTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATT ACAACCACA TTAATTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATT ACAACCACA TTAATTTC TAGATTGCAT CATATTTTAA ATATAACTAT CATATTTC TAGATTGCAT CATATTTC TAGATTGCAT CATATTTC TAGATTTC TAGATTGCAT CATATTTC TAGATTGCAT CATATTTC TAGATTCAT CATATTTC TAGATTCAT CATATTCAT CATATTCATATTCAT CATATTCAT CATATTCATATTCATATTCAT CATATTCAT CATATTCATATTCAT CATATTCATATTCAT CATATTCATATATTCATATTCATATTCATATTCATATTCATATTCATATTCATATTCATATTCATATATTCATATATTCATATTCATATATTCATATTCATATTCATATATTCATATTCATATATTCATATTCATATTC	18007
INGANATAT CAGTAGCTTG AATTAGACCA ATTAATTTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATCT ACAACCTGAA TTCTTTCTGT FINCANTTTG TCCRATTTTT TTCTCTAACA TTTATATCAC AAAGCAATTA ATTTGTGTGA TTTCTGCATA TGTATTTGTA ATTTCTTATA CATATTTGTA	18007 18107
INGAMATAT CAGTAGCTTG RATTAGACCA ATTAATTTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATCT ACAACCTGAA TTCTTTCTGT TCCANTITG TCCAATITTT TTCTCTAACA ITTATATCAC AAAGCAATTA ATTTGTGGA TTTCTGCATA TGTATTTGTA ATTCATCAG TCAATCAAT GTAGTAATAC TATATCATAA AATATACACA AATAATTGAG TGATAGGCTT CTAGTATAAG GACGGTAAGT TTGAAGGATG ATTCTATCTC GGCTGGCTAG	18007 18107 18207
IAGAMATAT CAGTAGETTG AATTAGACCA ATTAATTITE TAGATTGCAT CATATTITAA ATATAACTAT GTAATCATET ACAACCIGAA ITCITTETGT GTCAMITIG TECHNITITT TICTETACA ITTATATCAC AANGCAATTA ATTIGTGTGG TITTETGCATA TGTATITGTA ATTCATCAG TCAAATCAAT GTAGTAATAC IATATCATAA AATAATCACA AATAAATTGAG TGATAGGCTT CTAGTATAAG GACGGTAAGT TTGAAGCATG ATTICTATCTC GGCTGGCTAG TITACTCTGA GAAAGTTATT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCCAATG AACTGTTTTA TGTTCTGCTA	18007 18107 18207 18307
INGAMATAT CAGTAGGTTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATITIAA ATATAACTAT GTAATCATCT ACAACCIGAA ITCTITCTGT FIXCANTITG TCCAATITIT TICTCTAACA ITTAATCAC AAAGCAATTA ATTIGTGTGA TITCTGCATA IGTATTTGTA ATTCATCAG TCAAATCAAT SINGTAATAC TATATCATAA AATATCACA AATAATTGAG TGATAGGCTT CTAGTATAAG GACGGTAAGT TYCAAGCATG ATTCTATCTG GGCTGGCTAG HIACTCTGA GAAAGTTAT TITTATTGTT GGGTCTTAAG CTGACTTTAC ACACTTGGTG TCAGAATCAT TCCGGCAATG AACTGTTTA TGTTCTGCTA GGCTGATCAG CCAATCTAT ATGGCTGGTA ACAAAACAAT GTTTCCCAGT CATACCAACC ATGCCACCAT TITAACAGCT GATTAGGTA TYCAGAACTA	18007 18107 18207 18307 18407
IGAMMATAT CAGTAGCTTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATITHA ATATAACTAT GTAATCATCT ACAACCIGAA TICHTCIGT GTCAATTITG TCCAATTITT TICTCTACCA ITTATATCAC AAAGCAATTA ATTGTGTGA TITCTGCATA TGTATTGTA ATTCATCAG TCAATCAATG GTAGTAATAC TATATCATAA AATATACACA AATAATTGAG TGAAGGCTT CTAGTATAAG GACGGTAAGT TTCAAGGCATG ATTCTATCTG GGCTGCTAGG HIACTCTGA GAAAGTTATT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCCATG AACTGTTTA TGTCTGCTA GCCGATCGA CACACCTAT ATGGCTGTGA ACAAAACAAT GTTTCCCAGT CATACCAACC ATGCCACCAT TTTAACAGCT GATTAGTGTA TCCAGAACAT GTCCACCCCA TGTTCGTATG GCTGTTATTT AAAGAATGAAA GCAGTAGACA CACTTATTTT TTGAAAAATT TAGGCTCTGC AGGGCTCAATT ATTTTTTTTATA	18007 18107 18207 18307 18407 18507
INGAMATAT CAGTAGCTTG AATTAGACCA ATTAATTTTC TAGATTGCAT CATATITIAA ATATAACTAT GTAATCATCT ACAACCIGAA ITCATTCIGT FICANTITIG TCCAATITIT TICTCIAACA ITTATATCAC AAAGCAATTA ATTTGTGTGA TITCTGCATA IGTATTGTA ATTCATCAG TCAATCATC FINGTMATAC TATATCATAA AATATACACA AATAATTGAG TGAATAGGCTT CTAGTATAAG GACGGTAAGT TICAAGCATG ATTCTATCTG GGCTGCTAG ITTACTCTGA GAAAGTTATT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATCAT TCCGGCAATG AACTGTTTA TGTTCTGCTA GCTGATCAG CACAACCTAT ATGGCCTGGA ACAAAACAAT GTTTCCCAGT CATACCAACC ATGCCACCAT TITAACAGCT GATTAGTGTA TCAGAACAAT CTCCACTCCA TGTTCGTATG GCTGTTATCT AAAGATGAAA GCAGTAGACA CITTTATTATT TTGAAAAAATT TAGGCTCTGG GAGGGCAATT ATATTTGATA LUCAGGGGCT TITTTGAAG CAAACCAGAT ATATTTGTT AAGACTATCAT AAACCAGAACAAT LUCAGGGGCT TITTTGAAG CAAACCAGAT ATATTTGTT AAGACATTATCA AAGACGGAT ATCTTATTAA TTGGTACCATT AAATTGTGCA CACTTTCTCTC FINACTGTT CAGTACCTGT TCCAGCACTA ATACTGTCA AAGACATAAA CAAAGAACAA CACTTCATTAA TTGGTACCATT AAATTGTGCA CACTTCTCTCT FINACTGTT CAGTACCTGT TCCAGCACTA ATACTGTCCT AAGACATAAA CAAAGAACAA CACTTCATTAA CACTTCATTAA AGACTGTTATCAT AGACTGTTATCAT AGACTGTTATTAA CACTTCTCTTC GACGACCTAT ATACTGTCTCT GACGACCTT ATCTGTCTCT ATACTGTCT AAGACTGTTATTAA CACTTCATTCATTAAATTGTGCA CACTTCTCTCT AGACTCTCT ATACTGTCT AAGACTGTTATTATAA CACTTCTCT AGACTCTCT ATACTGTCT AGACTCT ATACTGTCT AGACTCTCT ATACTGTCT AGACTCT ATACTGTCT AGACTCT ATACTGTCT ATACTGTCT ATACTGTCT ATACTGTCT ATACTGTCT ATACTGTCT ATACTGTCT ATACTGTCT AGACTCT ATACTGTCT ATA	18007 18107 18207 18307 18407 18507 18607
IGAMMATA CAGTAGCTIC AATTAGACCA ATTAATTITC TAGATIGAT CATATITIAA ATATAACTAT GEAATCATCT ACAACCIGAA TICATTCIGT FICANTITIG TCCAATITIT TICTCTAACA ITTATATCAC AAAGCAATTA ATTIGTIGA TITCTGCATA IGTATIGTA ATTCATCAG TCCAATCAG FIGATATAC TATATCATAA AATATACACA AATAATTGAG TGAATGGCTT CTAGGTATAAG GACGGTAAGT TICAAGCATG ATTCTATCTG GGCTGCTAAG FITACTCTAG GAAAGTTATI TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCAGGCATG AACTGTTTA TGTCTGCTA GCCGATCAG CACATCTAT ATGCCTGTAA ACAAAACAAT GTTTCCCAGT CATACCAACC ATGCCACCAT TITTAACAGCT GATTAGTGTA TACAGACAT CTCCACTCCA TGTTCGTATG GCTGTTATTC AAACATGAAA GCAGTAGACA CITTATTTT TTGAAAAATT TAGAGCTCTGC AGGGTCAATT ATATTTGATA LINCAGGGGC TTTTTTGAAG CAAACTAGAT ATAATTTCTT TTGCAATTCT AAAGCCCTGAT ATCTTATTAA TTGCTACATT AAATTGTCA CCATTCTCT GTAACTGTTT CAGTACCTGT CTCAGGCCTGT ATCCTGGGGA AAAAAAAAAA	18007 18107 18207 18307 18407 18507 18607
INGALATAT CAGTAGCTTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATITIAA ATATAACTAT GTAATCATCT ACACCTGAA TICTITCIGT TICAATTITG TECAATTITG TECTATITTGT ATTATACACA TICAATTITGT ATTATACACA TICAATTITGT ATTATACACA TICAATTITGT ATTATACACA TICAATTITGT TECAATTITGT TICTAAGCATG TICAATCACACACAT TICAATCACACACACACACACACACACACACACACACACA	18007 18107 18207 18307 18407 18507 18607 18707 18807
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATGCAT CATATITAA ATATAACTAT GTAATCATCT ACACCIGAA ITCATTCTGT FINCANTITIG TCCAATTITI TICTCHACA ITTATATCAC AAAGCAATA ATTGGTGGA TICTCGCATA TGTATTGTA ATTCATCAG TCAATCATC FINGTAATAC TATATCATA AATATCACA AATAATTGAG TGAAGGCTT CTAGTATAAG GACGGTAAGT TTCAAGCATG ATTCATCTG GGCTGCCTAG TITACTCTAG GAAAGTTATI TITTATTGTI GGGTCTTAAG CTGACTTAC ACACTTGGTG TCAGAATCAT TTCAAGACGT AATCGTTAT TATTCTGCTA GCTCATCAG CCAAATCTAT ATGGCTGGAA ACAAAACAAJ GTTTCCCAGT CATACCAACC ATGCCACCAT TTTAACAGCT GATTAGGTA TCAGAACAT CTCACTCCA TGTTCGTATG GCTGTTATCT AAAGATGAAA GCAGTAGACA CTTTTATTTT TTGGAAAATT TAGGCTCTGG AGGGTCAATT ATATTTGATA LLIGAGGGGT TITTTTGAAG CAAACTAGAT ATALTTCTT TTGCAATTCTC AAAGCCTGAT ACTTTATTAA TTGGTACATT AAATCGTGCA CATTCTCCG GTAACTGTT CAGTACCTGT CTCAGGCACTA TACCAGGGGA AAAAATAAA GAAAGACCA GTGCCCAGAT AGCTTGGTCA GGGGCCCCTA ATCTTGCGG ACTAGTTC GGGAGCCCTA ATCTTGCGGG ACACAGGAA TTAAAGACAC ACACAGAA ATATAGAGTA TGGACTCGGA ACTCAGGGGT CTCACAGCCT TCAGAGCTGA GAGCCCCGAA CAGGATTA CCACATATT TATTGACAGC AAGCCAGTA TACATGTTAC TGAAACTATT CCTTATAGGA TTAAAAGGGA TGACTTGGCC TAGTTTATCTC CAGCAGGAAC ATGTCCTTAA GGCACAAATCA CATTATGCAA TCTCTCTGCG TTAAAGAACCA GTTTTCTGCC CTGGTGGGC CAGGTTTACTC CAGCAGGAAC ATGTCCTTAA GGCACAAATCA CATTATGCAA TCTCTCTGGG TTAAAACACCA GTTTTCTGCC CTGGTGGGC CAGGTTTACTC TCCACCCCTCAA ATGTCCTTAA GGCACAAATCA CATTATGCAA TCTCTCTGGG TTAAAACACCA CAGTTTATCTC CAGCAGGAAC ATGTCCTTAA GGCACAAATCA CATTATGCAA TCTCTCTGGG TTAAAACACCA CATTTTATCCC CTGGTGGGC CAGGTTTACCC CAGCCCCCAA	18007 18107 18207 18307 18407 18507 18607 18607 18807 18907
INGALATITA CAGTAGCTTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATITIAA ATTAACTAT GTAATCATCT ACAACCIGAA TICATTCIGT CICAATTITA TECTAACATT TICTCAACATTAGACA TICAATTCIGT TICAACTTTAGACA TICAATCACACACACACACACACACACACACACACACACA	18007 18107 18207 18307 18407 18407 18607 18707 18807 18907 19007
INGALATAT CAGTAGETTG AATTAGACCA ATTAATTTTC TAGATTGAT CATATITATA ATATAACTAT GEAATCATET ACAACCIGAA ITCITICIGE STOCKHITTI COCCAATTITI TICTCHACA ITTATATCAC AAAGCAATTA ATTIGGTGA TICTGCATA IGTATTGTA ATTCATCAG TCAATCAAT CAGTAGATTA TITATATCAC AAAGCAATTAA ATTIGGTGA TICTGCATA IGTATTGTA ATTCATCAG TCAATCAAT CAGTAGATCAT TICAACCATA AATATCACA AATAATCACA CAGTAGATCAT CAGATCAT TICAGACCAT TITAACAGCT GATTAGGTA TACGAACCAT CAGATCATA ATGCCTCGGAA ATTATTATA CAGATCAT CAGATCACA TITACAGCT GATTAGGTA TACAGAACAA CAGTAGACA ATCATTATTAT TACGACCAT TACACCAGC AGGGCCAATT ATATTTGATA LAGATGGAG CATTAGGAACAT CAGATCAGT CAGATCAGT CAGATCAGT AAACCTGCAGAT AAACTTGCAACAT CAGATCAGT AGGACACACAACTAA TACATTGCAA AAACAACAACAA ATATTCATA TACCAGCAGAA AAACAACAAA ATATTCAGAACAA CAGACAACAAA ATATTCAGACAA AAACAACAAA AAATAAAAAAAAAA	18007 18107 18207 18307 18407 18507 18507 18707 18907 19907 19107
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATGCAT CATATTTAA ATATAACTAT GTAATCATCT ACAACCTGAA TICTITCIGT FIXANTITIS TCCAATTTT TICTCAACA TITATATCAC AAAGCAATA ATTAGGTGA TITCITGCAT TITTIGGAA TITCIAGAT TICAAGCATG ATTCATCAG TAAACACA AATAATTGAG TIGAAGCAT CAGGTAAGT TICAAGCATG ATTCATCAG TAAACACA AATAATTGAG CAGGTTAAC ACACTTGGGT TCAGAAGCAT TICAAGCATG AATCATTTC GGCCCACAT TITAACAGCT GATTAGTAA TICTICIGCTA COCCACACCTAT ATTGATCATA AGAACAACAA GATTAGTAA GCAGTAGACA CACACCACAC	18007 18107 18207 18307 18407 18507 18507 18607 18907 19907 19107 19307
INGALATITA CAGTAGETTG AATTAGACCA ATTAATTITE TAGATTGAT CATATITIAA ATTAACTAT GEAATCATET ACAACCIGAA ITCAITCIGT FICANTITIS TECANTITIS TECCANTITIS TECCANTITIS TECCANTITIS TECCANTITIS TECCANTITIS TECCANTITIS TECCANTITIS TECCANTITIS TECCANTICAL ATTATACCA AATAATTGAG TGAATGAT CAACTTGGTG TCAGAATGAT TICLAGGCATG ATTCATACTG GEGGGTAGT TICLAGGCATG ATTCATACTG GEGGGTAGT TICLAGGCATG ATTCATACTG GEGGGTAGT TICLAGGCATG ATTCATACTG GEGGGTAGT GEGGATGAT ATTCATACTG GEGGGATGAT ATTCATACTG GEGGATGAT ATTCATACTG GEGGATGAT ATTCATACTA GEGGATGATG GEGGATGATT ATAATTCATCA GEGGATGATG GEGATGATGAT ATAATTCATCA AAGCCAGAT AAGCCAGATGAT ATAATTCATCA GAAGATGAT ATAATTCATCA GAAGAATAAA TACAGAGGA TACAGAGGAA TACAGAGAATA ACTAGAGAATA TACAGAGGA TACAGAGGAA TACAGAGAATA ACTAGAGAAACA TACAGAGGAA TACAGAGAACA TACAGAGGAA TACAGAGGAA TACAGAGGAA TACAGAGGAA TACAGAGGAA TACAGAGAA TACAGAGAACA TACAGAGGAA TACAGAGAA T	18007 18107 18207 18307 18407 18507 18607 18707 18907 18907 19107 19107 19407
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATGCAT CATATITAA ATTAACTAT GTAATCAT ACACCIGAA TICATCTCG FINGAMATA CAGTAGCTT TATCCAACA TITATATCAC AAAGCAATTA ATTAGGTGA TITCTGCATA FINGTMATAC TATATCATAA AATATCACA AATAATTGAG TGAATAGGCT CIAGTATAAG GACGGTAAGT TICAAGCATG ATTCATCAG FINGTMATAC TATATCATAA AATATCACA AATAATTGAG TGAATGGCT CIAGTATAAG GACGGTAAGT TICAAGCATG ATTCATCAG CAGATCATAT FINGACAGA GAAAGTTAT TITTATTGT GGGTCTTAAG CIGAGTTAC CAACCTAGTG TCAGAACAT TICAAGCATG ATCCTTACTA TGCTCGCTAG COCACTCCA TGTTCGTATG GCTGTTATC AAAACAAT GTTTCCCAGT CATACCAACCA TGCACCACT TITAACAGCT GATTAGGTA TCAGAACAT LUCAGGGGC TITTTTGAAG CAAACTAGAT ATAATTGTT TIGCAATTCT TAGAACAATAAA GAAAGAACAA GTGCCACAAT LUCAGGGGC TITTTTGAAG CAAACTAGAT AAAATTATTCTT TIGCAATTCT TAGAACATATA TAGGCTCTGC GAGGCCCATA AACTGTGCG CACAACCTA TACCAGGGAG AAAAAAAAA GAAAGAACAA GTGCCACAACT ACCTTGGCAG GAGGCCCTA AACTGTGCG CACAACCTA TACCAGGAGA AAAAAAAAA GAAAGAACAA GTGCCACAACT ACCTTGGCAG GAGGCCCTA AACTGTGCG CACAACTAT TAATGACAC AAGCCAGAA AACAAGAGA TGGACTGGA AATAAAGGGA TGACTCTGGC CAGGGCCCGAA CAGACGATTA CACACAACAT TAATAGCAC AAGCCAGAA AACAAGAGA TGGACATAACA CCTTTAAGCA GTTTCCCCC CAGGGTGGGC CAGGGTTTCC TAGCCCCCAAC CACACCTT CCAGTGGGA TACCAGGAGC TACCAGAGC TACCAGAGC TAACACTATA TACCAGGAGC TACCAGAGAA TACCAGGAGA TACCAGGAGA TACCAGGAGA TACCAGGAGA TACCAGAGAA TACCAGGAGA TACCAGGAGA TACCAGGAGA TACCAGGAGA TACCAGGAGA TACCAGGAG TACCACACCTT CCAGGAGC TACCACACACTT CCAGGAGC TACCACACACTT CCAGGAGC TACCACACACTT CCAGGAGC TACCACACACTT CCAGGAGC TACCACACACACT TACCAGGAGC TACCACACACACACACACACACACACACACACACACAC	18007 18107 18207 18307 18407 18507 18507 18607 18907 19907 19107 19207 19307 19307
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATTGAT CATATITAA ATTAACTAT GTAATCATC ACACCIGAA TICATTCIGT CICAMITTO TCCAMITTO TCCAMITAGA ATTATACACA ATTATACACACAC	18007 18107 18207 18307 18407 18507 18607 18707 18907 18907 19107 19107 19407
INGALATAT CAGTAGETG AATTAGACCA ATTAATTTC TAGATGCAT CATATITAA ATTAACTAT GEAATCATE ACAACGGAA TICTITCIGT STANTING TCCAATITIT TICTCAAACAT TITAATCAC AAAGCAATTA ATTIGGTGA TICTCGCATA GEATCATCT ACAACGGA TACAACATA ATTAATCAC AAACAATTA ATTIGGTGA TICTCGCATA GEATCATCT ATTACAGG TACAACATA ATTAATCACA AATAATCAC AATAATCAC TAGATGCT CAGGATGAT TICTAGCAC ATTICATCT GEGEGATG TICAACACA AATAATCAC CACAACCATA AAGCAATCATA ATGCCCCGGAA ACAACCATA CACATCGGG TACAACACATA CACATCGGG TACAACACATA ATTACAGACAT CACAACCATA AAGCACACATA AAACACATA GETACCACAC AAACACATA CACATCGGG AAGAACATA TICAACACAC AAGCACCATA TACAACCATA AAACACATA AAACACATA AAACACACAC	18007 18107 18207 18307 18407 18507 18507 18607 18907 19107 19107 19207 19507 19507 19507 19707
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATGCAT CATATTTAA ATTAACTAT GTAATCAT ACAACCTGAA TICATCTGT CACACTTTA ATTAACTAT GTAATCAT ATTACACTAT TICTCAATTTT TICTCAACA TITATACCA AAAGCAATTA ATTAGGTGA TITCTGCATA TICAAGCATA AATTACACA AATAATTGAG TGAATGGCT CTAGTATAAG GACGGTAAGT TICAAGCATG ATTCATCAG TAATCATAA AATTACACA AATAATTGAG TGAATGGCT CTAGTATAAG GACGGTAAGT TICAAGCATG ATTCATCAG TAATCATAA AATTACACA AATAATTGAG CTGAGTTAC CAACCTAGAACAT TAATAGCATA TAGGCTCACACAT TITTAACACAT TICAAGCATA ATTCATCAG CTCACCACAT TITTAACACAT ATTCATCAGAACAT CTCACCACACATTA TAGGCTCAGAACAAACAAA GAAACAAA GAACCAACCAACATAAA AACACTAGAAA GAACCAACCAAACAAAA AACACAGAAA AACACTAGAAA GAACCAACCAAACAAAAAAA GAACCAACCAAAACAAA AACACAGAAA AATAAAAAAAA GAACCAACCAAA AACACAGAA AACACAGAA AACACAGAA AACACAGAAA AACACAGAAA AACACAGAAA AACACAGAAA AACACAGAAA AACACAGAAA AACACAGAAA AACACAGAAAA CAAACAAA	18007 18107 18207 18307 18407 18507 18507 18507 18907 19107 19207 19307 19507 19507 19507 19507 19507
INGAMATA CAGTAGETG AATTAGACCA ATTAATTTC TAGATGACA CATATTTAA ATTAACTA GEAATCAT ACAACCTGAA TICTITCIGT STANTING TCCAATTTT TICTCAACA TITAATCAC AAAGCAATA ATTGGGGA TICTIGGAA GEAGGATA TICTIGGA TICAACTTT TICTCAACA TATAATCAC AAACAACTA ATTAGGGT CAGGATAGT TICTIGGAA GEAGGATAGT TICAACACT ATTACACAC AATAATCAC AAAACAACTA CACATTGGGG CAGGATAGT TICAACACT AAACACTAC AACACTTAGA CACATTGGGG CAGAACTAA TICTITCIGTA GCCAATCTAA ATTAGACTA AACACTTAGAA CACATTGGGG CAGAACTAA TICTITCIGTA GCCAATCTAA AAGAACAACTA AAACACTAA GCCACCAT TATAACAGC GATTACTGTA TACAGACAACTA CACATTGGGG CATACACTA TACACTGGAACACT CACATTGGAACACT CACATTGGAACACACT CACATTGGAACACT CACATTGGAACACACT CACATTGGAACACT CACATTGGAACACACT CACATTGGAACACACT CACATTGGAACACT CACATTGAACACT CACATTACACACT CACATTACACT CACA	18007 18107 18207 18307 18407 18507 18507 18707 18907 19107 19107 19107 19407 19507 19507 19507 19807 19807
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATGCAT CATATTTAA ATTAACTAT GTAATCAT ACAACCTGAA TICTITCIGT TACAATTTT TICTCAAACAT TATATCACA AAAGAATTA ATTAGGTGA TICTCGCATA TICTAGAT TATATCACA AATAATTGGA TGATAGGCTT CTAGTATAAG CACCTGGATAGT TICTAGAGA ATTACACTA GATACACTA GATACACTA AATAATCACA AATAATTGGA TGATAGGCTT CTAGTATAAG CACCTTGGTG TCAGAACATA TICTACACAC CACAACCTAT ATGACGCTGGAAAG CACAACCATA ATGACTGCTAAG CTAGAACACAT CACAACCATA ATGACTGCTAAG CACAACCATA ATAATTGATA ACACATGGTG TCAGAACACAT TATAACACCA GATTACTGTA TICTAGAACAC COCACACCACA TITTAACAGCA GATTACTGTA TICAGAACACA CACAACCATA ATAATTGATA AAAACAAAAAAAAAA	18007 18107 18207 18307 18407 18507 18507 18607 19907 19107 19207 19407 19507 19507 19507 19507 19907 20007 20107
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATGCAT CATATTTAA ATTAACTAT GTAATCAT ACAACCTGAA TICATCTGT FINGALATAC TATATCATAA AATTACACA AITAATTGC TAGATGCAT CINGTAINAG GACGGTAAGT FINGALATAC TATATCATAA AATTACACA AATAATTGGG TGATAGGCTT CINGTAINAG GACGGTAAGT FINGALATAC TATATCATAA AATATACACA AATAATTGGG TGATAGGCTT CINGTAINAG GACGGTAAGT FINGALAGACAACTATA ATTATCATAA AATATACACA AATAATTGGG TGATATAC CACATTGGGT TCAGAACAT FINGALAGACA AAGATTATT TITTATTGTT GGGTCTTAAG CTGAGTTAC CALACCTAC ATGCCACCAT FITTAGGACTGA ACCAAACTATA AGAATGATA GACAGAGAA CACTTGGGT AAGACTATAT TATATCACAC FINGALGGGCC TITTATGGAG CAAACTAGAT AAAATTGTT TAGAAACAT TAGATTGATA GAAACAAA GAAACAAA GAAACAAA GAAACAAA ATAATAGGCC FINATGGTT CAGTACCTGT CTCAGACCATA AAAACAATAAA GAAACAAA GAAACAAA GAAACAAA ATAATAGGCA TAGATTGATAA TAGACTGCAGAC MIMAGAGGAA TTAAAGACA AACCACAGAA ATAATAGAGAT TAGACTATAC CITTAAGCA GATCACGCCT MIMAGAGGAA TTAAAGACA AACCACAGAA ATAATAGAGAT TAGACTATAC CITTAAGCA GATCACGCCT MIMAGAGAA TAGAATTAC CACACAGAA AACCAGAGAC TAGACTATAC CITTAAGCA GATCACGCCC MIMAGAGAA TAGAATTACA AACCAGAGAC TAGAATTAC TAGAACAAA CITTAACAC CACACCAT CACACCACACACACACACACACA	18007 18107 18207 18307 18407 18507 18507 18507 18907 19107 19207 19307 19507 19507 19507 19507 19707 19507 19707 19807 20007 20107
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATTGCAT CATATTTAA ATTAACTAT GTAATCAT ACAACTGAA TTCTTCTGT FINGATATA CACATITAT TITCTCAACA ITTAATACAC AAAGAATA ATTGTGTGA TTCTGCATA TTCTTGTA ATTATCAG TCAATCAT FINGATATAC TATACCATA AATATCACA AATATTGGG TGAATTACA CACATTGGTG TCAGAATCAT FINGATCAG CACAATCATA ATTGTTTGTG GGGTCTTAAG CTGACTTTAC CACATTGGTG TCAGAATCAT FINGATCAG CACAATCATA ATGGCTGGTAAG CTGACTTTAC CACATTGGTG TCAGAATCAT FINGATCAG CACAATCATA ATGGCTGGTAAG GTTATGATAA GCACATCGTG CATACCACC ATGCCACCAT FINALACAT ATGGCTGTAG GCTGTATCT AAAGATGAAA GCAGTAGACA CACTTGGTG TCAGAATCAT FINALACAT GGTCGTATA GCAGACTAGAT FINALACAT ATATTGAAG CAAACTAGAT FINALACAT TATACAGCA ATATTCTT TTCAATCACAC AAACCACAGAA FINALACAT TATACAGCA ATATTCTT TTCAATCACAC ACCACAATCAT TATACAGCA CACACCAGAA FINALACAT TATAGACAC CACACCAGAA FINALACACA ATATTGGAGA FINALACAC AGCACAGAA FINALACAC FINALACAC AGCACAGAA FINALACAC AGCACAGAA FINALACAC AGCACAGAA FINALACAC FINALACAC AGCACAGAA FINALACAC FINALACAC FINALACAC AGCACAGAA FINALACAC FINALAC	18007 18107 18207 18307 18407 18507 18507 18507 18907 19107 19207 19307 19507 19507 19507 19507 19707 19507 19707 19807 20007 20107
INGALATAT CAGTAGCTG ANTACACA ATTACTOCT CAGATGCAT CATATITAA ATATACTAT GEARTCATC ACACCIGAA ITCITICIST CICCANTITI TICCTAACA ITLATACAC ANAGGANTA AITGTGGA TITCIGCATA TGATTIGTA ATTACACA CANACANT STATTCACA TATACACAC ANAGANTA AITGTGGAT CAGATGAT CAGAGGAT AITCITICIST GEOGGATAGAT TATACACAA ANTACACA ANTACACTAGG CAGAGGATACAT TICAGGATGAT TICAGGATGAT CAGACTAGTAT ATGCTCACAC AITGGGATGAT CAGAATGAT TICTGGATA AGGCGCTAGATTACA AGGCGCTAGATGAT ATGCTCACTGA ACAAACAAT GATTACCACC ATGCCACCAC ATGCCACCAC TATACACAC AGGGCACATTACACACCAC AGGCCCCACACCATACACAC AGGCCCCCACACCATACACACAC AGGCCCCCACACCACA	18007 18107 18207 18307 18407 18507 18507 18507 18907 19107 19207 19307 19507 19507 19507 19507 19707 19507 19707 19807 20007 20107
INGALATAT CAGATACCTA ANTAGACCA ATTAATTCC TAGATACCAT CATATTTAA ATTAACATA GEALACATCA ACACCAGA TECHTICISE SECONATITY TECCHANTITY TECCHACA TATAATCAC ALAGCANTA ANTAGAGC TECCHATITY TECCHACA TATAATCAC ALAGCANTA ANTAGAGCA TATAACCAC ACACACCATA TATAACCAC ALAGCANTAA ANTATACCA ALAGCANTAA ANTATACCA ALAGCANTAA ANTATACCA ALAGCANTAA ANTATACCA ALAGCANTAA CACAACCACA TECCHACTATI TITAATTCT GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GETTACCACAC ACACTACTA TATAACCAC ACACACTAA TAGGCTEGA ACAAACAAA GTTACCACACCA TAGCCACCAT TATAACCAC AGGGTCAAT ACACACACAA ACACACAAA ACACACACAA ALAGCATA ATAATTCTT TICAATTCT TAGAACACA ACCCACACA TAGAACCAT TAGACCACAA TATAACACAC CACACACAAA ATAATACACAT TACCAACACAA ACACACAGA ATAATACACAT TACCAACACAA ATAATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC CACACACACAA ACACACAGA ATATACACAC CACACACACAA ACACACACAAA ATATACACAC TACAACACAC ACACACAC	18007 18107 18207 18307 18407 18507 18507 18507 18907 19107 19207 19307 19507 19507 19507 19507 19707 19507 19707 19807 20007 20107
INGALATAT CAGTAGETT AATTACACA ATTAATTTC TAGATICGAT CATATITATA ATTAACATA GRANTITAT CACATICTAT TACCATACA TITAATACA ATTAATACA ATTAATACA ATTAATACA ATTAATACA ATTAATACA CAAAGAATAA ATTATAGAGT TAGATAGAGT TAGATAGA	18007 18107 18207 18307 18407 18507 18507 18507 18907 19107 19207 19307 19507 19507 19507 19707 19807 20007 20107 20207
INGAMATAT CAGTAGGTTG MATTAGACCA ATTAATATIC TAGATIGGAT CATATITTAA ATATACAT GRAATCAT ACAACCIGAA TICITICIST GROWANTIG TOCKANTITI TICTCHAACA TITAATACA AAGGAATTA ATTIGTGGA TITCTGCATA GEATTIGTA ATTACACAAG TOLAATCAAT GRAATATA AATATCACA AATATTCAC AAGGAATTA ATTIGTGGA TITCTGCATA TICAGAGGATG ATTCATACAG GOCTGGCTAG GRAGATTATI TITTATTGTT GGGTCTTAAG CRAGTAGAT CAGATAGAT TICAGAGTAGAT TICAGAGGATG GATAGTTAT ATGGCTGTAA ACAACCACAT CAGATCAACCAT TITAACAGCT GATTACTGTA TICAGACCAT COCACTCCA TGTTCGTATG GCTGTATCT AAGATGAAA GCAGTAGACA CATTIATTTT TIGAAAAATT TAGGCTCTCC AGGGTCAATT ATATTTGATA GACACCACA TGTTCGTATG GCTGTATCT AACAGGGAGA ACAAATAAA GCAGTAGACA TAGCCTGAT ATATTTGATA TAGCAGCATA AAATTTGATCA GACACTCCA TGTTCGTATG GCTGTATC AACACCACACAA AATACAGGAA AACACCACACACACACACACACACACACACAC	18007 18107 18207 18307 18407 18507 18607 18707 18907 19007 19107 19107 19407 19507 20107
INGALATAT CAGATACCTA ANTAGACCA ATTAATTCC TAGATACCAT CATATTTAA ATTAACATA GEALACATCA ACACCAGA TECHTICISE SECONATITY TECCHANTITY TECCHACA TATAATCAC ALAGCANTA ANTAGAGC TECCHATITY TECCHACA TATAATCAC ALAGCANTA ANTAGAGCA TATAACCAC ACACACCATA TATAACCAC ALAGCANTAA ANTATACCA ALAGCANTAA ANTATACCA ALAGCANTAA ANTATACCA ALAGCANTAA ANTATACCA ALAGCANTAA CACAACCACA TECCHACTATI TITAATTCT GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GEGETATAA GETTACCACAC ACACTACTA TATAACCAC ACACACTAA TAGGCTEGA ACAAACAAA GTTACCACACCA TAGCCACCAT TATAACCAC AGGGTCAAT ACACACACAA ACACACAAA ACACACACAA ALAGCATA ATAATTCTT TICAATTCT TAGAACACA ACCCACACA TAGAACCAT TAGACCACAA TATAACACAC CACACACAAA ATAATACACAT TACCAACACAA ACACACAGA ATAATACACAT TACCAACACAA ATAATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC ACACACAGA ATATACACAC CACACACACAA ACACACAGA ATATACACAC CACACACACAA ACACACACAAA ATATACACAC TACAACACAC ACACACAC	18007 18107 18207 18307 18407 18507 18507 18507 18907 19107 19207 19307 19507 19507 19507 19707 19807 20007 20107 20207
INGLANATA CAGTACCTIC ANTRACACA ATTANTITIC TAGATTGCA CATATITICA ATTANCAT GTALTCATC ACACCICAN ITCITICACE GOOGNATA ATTACACA ALACACATA ANTRACACA ALACACATA ANTRACACA ALACACACACA ATTACACACA ALACACACACACACACACACACACACACACAC	18007 18107 18207 18307 18407 18507 18607 18707 18907 19007 19107 19107 19407 19507 20107
INGALATAT CAGTACTTC ANTRACACA ATTANTITIC TAGATTCAT CATATTITIA ATATACAT GTAATCATC ACACCICLA TICTITICE TOCKNITTY TICTICANA ANTATACACA ANTANTICA CAMAGGATTA ATTITICAGATA TOTATCAGA TATATCATA ANTATCAGA TAGATCATA ATTATCAGA CAGATCATA TAGATCATA ATTATCAGA CAGATTATA CACATTGGAT CACATCAGA TACATCAGA TACATCAGA CACATCATA ATTATCATAGA CACATCATA ATTATCAGA CACATCATA ATTATCAGA ACCATCAGA CACATCAGA CACATCATA ATTATCAGA CACATCAGA TAGATTACA TAGATTCATA TAGATCAGA CACATCAGA TAGATTACA TAGATTCAGA TAGATCAGA TAGATTACA TAGATTACA TAGATCAGA TAGAT	18007 18107 18207 18307 18407 18507 18507 18707 18707 19007 19107 19107 19507 19507 19507 19507 19507 19507 19707 19707 19707 19707 19707 20007 20107 20307
INGALATAT CAGTACCTIC MATRACICA ATTATTTC TAGATTGCAT CATATITAA ATATACCTAT GRATTCATCT ACCACCAGA TACTTCAGE CAGATITATITIC CORATITITI TICCTARCA TITATACCAA AAGAATAA ATTATTCAGA GACAGATAA ATTATTCAGA TACATTCAGE CAGATCATITIC CAGATCAGE CAGATCATITIC CAGATCAGE CAGATCAGE CAGATCAGE CAGATCAGE CAGATCAGE CAGATCAGE CAGATCAGE CAGATCAGE ATTATACCAGE CAGATCAGE CAGATCAGE CAGATCAGE CAGATCAGE ATTACCAGE CAGATCAGE CAGATCAGE ATTACCAGE CAGATCAGE CAGATCAGE ATTACCAGE CAGATCAGE CAGATCAGE ATTACCAGE CAGATCAGE ACCATACCAGE CAGATCAGE CAGATCAGE ACCATACCAGE CAGATCAGE ATTACCAGE CAGACCAGE ACCATACCAGE CAGACCAGE CAGACCAGAGACA CACACCAGE CAGACCAGAGACA CACACCAGE CAGACCAGAC	18007 18107 18207 18307 18407 18507 18607 18707 18907 19007 19107 19107 19407 19507 20107
INGALATAT CAGTACCTC MATRACCCA ATTATTTC TAGATTGCAT CATATITAA ATATACCTAT GENATCATC ACACCTGAA TTCTTCGTCATA TITTATTCATCA ATTCATACA ANACAMTA ATTATTCAGATA TOTATTCAGATA ATTCATACAG CANACIANA CHARTATTCA CANACCAMATIA ATTATCAGATA ATTCATACAG CANACIANA CHARTATTCA CANACCAMATIA ATTATCACA ATTATTCAC CANACCAMATIA ATTACCAGATA ATTATTCAGATAGAT CHARTATCAGATAGATATACAGATAGATATACAGATAGATATACAGATAGAT	18007 18107 18207 18307 18407 18507 18507 18707 18707 19007 19107 19107 19507 19507 19507 19507 19507 19507 19707 19707 19707 19707 19707 20007 20107 20307
IGGAMATAT CAGATACCTTC ANTIGATICA ATTANTITIC TAGATGCAT CATATITITAA ATTAACTAT GTAATCATC ACACCIGAA TICTITICISCE FICCANTITIC TICCANTITIT TICTCANACA ITTAATATCA ANAGCANTA ATTAGTGTA TICTICCATA TICATATATATA ANTIATACCA ANTIGATICA CAATGATTA TATACTATA ANTIATACCA ANTIGATICA CAATGATTA TATACTACA ANTIATACCA ANTIGATICA CAATGATTA CACATGATA TICACACA GEGGGATA ATTATATATA ANTIATACCA GEGGGATACT TITAACAGCT GATTACTACA GEGGGATACT TITATATACA GEGGGATACT ATTATATATA ANTIATACTAC GEGGGATACT TITAACAGCT GATTACTACA GEGGGATACT ATTATATATA ATTACCACCACA ATTACCACCACA TITAACAGCT GATTACTACA TITACAGCT GATTACTACACACACACAT ANAGACAAA ACACACACA ATAATTACTATA GAACCACACACACAT TATACAGCT GATTACTACACACACACACACACACACACACACACACACA	18007 18107 18207 18307 18407 18507 18507 18707 18707 19907 19107 19107 19407 19507 19507 19507 19507 19507 19507 19507 20107 20207 20307
IGGALITIT GECCANTITY TICCTANCA TITALATICA ANGCANTA ATTITATA ATATACTA GENATICATE ACALCAGA TICTITAGA GENATITATA TITALATACA ANGCANTA ATTITAGA TICTICAGA TICTITAGA TICTICAGA TICTITAGA TICTICAGA ATTITAGA GENATACA TICTICAGA ANGCANTA ATTITAGA GENATACA ATTITATATO GENATACA CALAGACATA ATTICAGA TICTICAGA ANACATATA ATTIGAGA GENATACA CALAGACATA CACATAGA TICCAGACA TICCAGACA ATTITATORIC GENATACA GENATACAGA CACATAGA TICAGACATA ATTICAGACA GENATACAGA CACATAGA ATTICAGACA CACATAGA CACATAGA TICAGACATA ATTICAGACA CACATAGA ATTICAGACA CACATAGA ATTICAGACA CACATAGA ATTICAGACA CACATAGA ATTICAGACA ATTICAGACA CACATAGA ATTICAGACA ATTICAGACACA ATTICAGACACATA CACACAGAA ATTICAGACACA ATTICAGACACA ATTICAGACACA ATTICAGACACATA CACACAGAA ATTICAGACACA ATTICAGACACA ATTICACACACA ATTICACACACACACACACACACACACACACACACACACA	18007 18107 18207 18307 18407 18507 18507 18707 18707 19907 19107 19107 19407 19507 19507 19507 19507 19507 19507 19507 20107 20207 20307
REGALLITE CCALITITE TICCTIANCA ITTAINATICA ANGCANTIA ATTITATA ATATACTAT GRANICATE ACAGCICIAA TICTITICITE TICCATITITI TICCTIANCA ITTAINATICA ANGCANTIA ATTITICATICA TICTICATA ITTAINATACA ANGCANTIA ATTITICATAGA TICTICAGA INGTAITTE TICATAGA ATTAINATICA CANACATAGA ATTITIATATACA ANGCANTIA ATTITICATICA GEOGRAFIA ATTITICA ANGCANTIA TICTIATATACA ANGCANTIA CANACATAGAT CANACAGA TICCAGCAN CANCULAR ATTICATATA GEOGRAFIA GEOGRAFIA ATTITICATAGA GEOGRAFIA ATTICAGA CANCULAR ATTICAGA GEOGRAFIA ATTITICATA CANCULARCA ATTICAGA TICCAGACA CANCULAR ATTICAGACA CANCULAR ATTICAGACA CANCULAR ATTICAGACA ATTITICATAGA ANGCATAGA ATTITICATAGA CANCULAR ATTICAGACA ATTITICATAGA ANGCATAGA ATTITICATAGA ANGCATAGA ATTITICATAGA ANGCATAGATA ATTITICATA ANGCACACA ATTITICAGACA ATTITICATAGA ANGCATAGA ATTITICATA ANACACAGA ATTITICAGACAT ANGATICAGA ATTITICAGACA ANGCAGAGA ATTITICAGACA ANGCAGAGA ATTITICAGACA ANGAGATAAA CANACACAGA ATTITICAGACAT ANGAGATAAA CANACACAGA ATTITICAGACATA ANACACAGA ATTITICAGACA ANGAGATAAA CANACACAGA ATTITICAGACA ANGAGATAAA CANACACAGAA ATTITICAGACA ANGAGATAAA CANACACAGA ATTITICAGACA ANGAGATAAA CANACACAGA ATTITICAGACA ANGAGATAAA CANACACAGA ATTITICAGACA ANGAGATAAA CANACACACACA ANGAGAACA CANACACAGACA ANACACAGA ATTICAGACACA ANGAGAACAA CANACACAGA ATTICAGACACA ANGAGAACAA CANACACAGA ATTICAGACACA ANGAGAACAA CANACACAGA ATTICAGACACA ANGACACACACACACACACACACACACACACACACACACA	18007 18107 18107 18207 18307 18407 18507 18507 18707 19907 19107 19107 19507 19507 19507 19507 19507 19507 19507 20007 20107
HADMATAT CAGTAGCTTC ANTTAGACCA ATTANTITIC TAGATISCAI CATATITATA ATATACTAI CHANTCAIC ACACCIGAA TRICTICATE INCOMINE OCCURRING TOTAL TRICTICA TRAINCAI CHANACAT TRAINCAICA TATACACA ANTANTACA CAANCANT ANTISTICAGA TATACACATA ANTATACACA ANTANTACAG CAAGCATAA ANTATACACA ANTANTACAG CACATICAT TATACATAA ANTATACACA ANTANTACAG CACATICAT TATACACAC ANTACAT CONCENTRAC CACATICAT TATACACAC ANTACAT CONCENTRAC CACATICAT TATACACAC ANTACATACAC CACATICAT TATACACAC ANTACACAC CACATICAT THACACAC THITTICAGA CACATICAT ANTACACAC ANTACACACAC ANTACACACACACACACACACACACACACACACACACACA	18007 18107 18207 18307 18407 18507 18507 18507 18707 18907 19107 19207 19307 19507 19507 19507 19507 19507 19507 20107
HADMATAT CAGTAGCTIC ANTRACACA ATTANTITIC TAGATICGAI CATATITHA ATANACTAI CHARTCAIC ACACCICAL TICTITICAGE FINANTITY TICCANTITY TICTACACA TITATATICA ANAGGANTA ATTITICAGAI TATATICATA TATATICATA ATTANACATA CHARTATICAGE TATATICAGA TATATICAGA TATATICAGA TATATICAGA TATATICAGA TATATICAGA ATTANACATA CHARTATICAGA CACATICATA CHARTANAC CACATICATA TATATICATA ATTANACATA ATTANACATA ATTANACATA ATTANACATA ATTANACATA ATTANACATA ATTANACATA ATTANACATA CACATICAGA CACATICATA TACAGACATA TACAGACATA ATTANACATA ATTANACATA ACACATICATA TACAGACATA TACAGACATA ATTANACATA CACATICATA TACAGACATA ATTANACACA CATACACACA ATTANACACA TACAGACATA ATTANACACA CACACACATA ATTANACACA ATTANACACA TACAGACATA ATTANACACA CACACACATA ATTANACACA ATTANACACA ACCAGACA ATTANACACA ACCAGACA ATTANACACA ACCAGACA ATTANACACA CACACACACA ATTANACACA ACCAGACA ATTANACACA ACCAGACA ATTANACACA ACCAGACA ATTANACACA ACCAGACA ATTANACACA ACCAGACA ATTANACACA ATTANACACA ACCAGACA ATTANACACA ACCAGACA ATTANACACA ATTANACACA ATTANACACA TACACACACA TACACACACA TACACACACA ATTANACACA TACACACACA TACACACACA TACACACACA TACACACAC	18007 18107 18207 18307 18407 18507 18507 18707 18707 19907 19107 19107 19507 19507 19507 19507 19507 19707 19707 19707 19707 19707 20007 20107 20307 20478 20559 20657 20957
IMAMATA CATACCTTC ANTACACCA ATTACTTC TAGATTCCA CATATTTA ATATACACT GLARICAGA TACTTCTCTC INCANTIFI TECCATATIT TICCTARA TITATACAC ANGCARTA ATTITICACATA GITATTCTA ATTACCACA ANTACACAT INGINIAL TATACACA ANTACACA ANTACTACA CATACACT INGINIAC TATACACAA ANTACACA ANTACTACA INGINIAC CACACACTAT ATCOCCACA ANTACACAT CONCINCA GAAAGTATAT TITTACTAC GACCTTAAAC CACACTCORY CONCINCA GAAAGTATAT TITTACTAC CACACTCORY CONCINCA GAAACTATA TIGGETCOA ACAAACAAT CONCINCAC CACACACTAT ATCOCCACA ACAAACAAT CONCINCAC CACACACTAT GACCTATACT ACAACACAT CONCINCAC TOTACCATA CACACACAA ATAATTCTT TICCATACT CONCINCAC TOTACCATA CACACACAA ATAATTCTT TICCATACT CONCINCAC TOTACCATA CACACACAA ATAATTCTT TICCATACT CONCINCAC TOTACCATA CACACACAA ATAATTCT TICCATACT CONCINCAC TATACCAA TAAATTCCAT CONCINCAC CONCINCACACACACACAA ATAAACACAT TICCATACAC CONCINCACACACACACACAA ATAAACACAT TICCATACAC CONCINCACACACACACACACAA ATAAACACAT TICCATACAC CONCINCACACACACACACACACAA ATAAACACATA CONCINCACACACACACACACACACACACACACACACACACA	18007 18107 18107 18207 18307 18507 18507 18507 18707 19007 19107 19107 19507 19507 19507 19507 19507 19507 20007 20107
HAMMATA CARTACTTC ANTACACCA ATTACTTC TAGATTCCA CATATITTA ATTACACTA GLACCTGA TICTTCCTATA TO CCANTITT TICCTATACA TITATACCA ANGCARTA ATTACTCA ANTATACACA ANTATACACA ANGCARTA ATTOCTATA GLACATICAT GLACATICAT GLACATICAT GLACATICAT TITATACCACA ANGALATICAT CAGATTACA CACTICATE GLACATICAT TICACACAC ATCACACACACACACACACACACACACACACA	18007 18107 18107 18207 18307 18407 18507 18507 18707 19907 19107 19107 19507 19507 19507 19507 19507 19507 19507 2007 20107 2
INGIAINATA CAGTACCTT ANTAGACA ATTANTITC TAGATICAT CITATITIAA ATTANACTAT GENATCATC GALACTIGA TICTITCETE ENCINTING TECHNITIT TICTICACA TITANACATHA ATTANACAC ANTANTICAC ANTAGATATA TITTOTICACATA TITATICACAA TITANACACA ATTANTICAGA TITATICACAA TITATICACAACAA TITATICACAACAAA TITATICACAACAAA TITATICACAA TITATICACAACAAA TITATICACAAAAAAAAAA	18007 18107 18107 18207 18307 18407 18507 18507 18707 19907 19107 19107 19507 19507 19507 19507 19507 19707 20007 20107 20207 20307
INGIAINATA CAGTACCTT ANTAGACCA ATTANTITC TAGATICGA CATATITATA ATTANACTA GRANCATTA ATTACTACA GRANCATTA ATTACACA GRANCACTA ATTATTATA ATTACACA GRANCACTA ATTATTATA ATTACACA GRANCATTA TITATTATA ATTACACA GRANCATTA TITATTATA ATTACACA GRANCATTA TITATTATA ATTACACA GRANCATTA TITATTATA ATTACACA GRANCATTATA TITATTATA GRANCATTA TITATTATA GRANCATA TITATTATA GRANCATA TITATTATA GRANCATA GRAN	18007 18107 18107 18207 18307 18507 18507 18507 18707 19707 19107 19107 19107 19507 19507 19507 19507 19507 19507 20107 20207 20107 20307 20478 20559 20657 20757 20957 21157 21157
INGIAINATA CAGTACCTT ANTAGACA ATTANTITC TAGATICAT CITATITIAA ATTANACTAT GENATCATC GALACTIGA TICTITCETE ENCINTING TECHNITIT TICTICACA TITANACATHA ATTANACAC ANTANTICAC ANTAGATATA TITTOTICACATA TITATICACAA TITANACACA ATTANTICAGA TITATICACAA TITATICACAACAA TITATICACAACAAA TITATICACAACAAA TITATICACAA TITATICACAACAAA TITATICACAAAAAAAAAA	18007 18107 18107 18207 18307 18407 18507 18507 18707 19907 19107 19107 19507 19507 19507 19507 19507 19707 20007 20107 20207 20307

# Figure 8D

ACTICAGECT	GAGTGACAGA	GTAAGACCCT	ATCTCAAAAA	ACAGAAAAA	AAAAACACTG	GCCCAAAGGA	AATGAACTTG	TTACAGAAGC	CGGGGTTTAN	••: 1
AACACCAAA1	ANTGENETIG	TACCTAGTCC	TTCCCGGGTG	CTCTGCAGAC	ATTTCTCCAA	GCGTAGTCTG	CALACAACCT	ACATATGTAG	AATTACCE	21:
ARRACCCAC	CHARMOCCA	CCCCCCCCC		G. G. C.		ACCTUCAGE T	MAN OCCURATE	ACATICCGTC	TGTGAGARGA	21e 217
ATTTALLT	ALLTTARCEL	CALTRALICAC	A		CICHINITAN	OLIGIACIA!	COMINGGIA	VCCVCTVCCC	ACATATICCO	218
CTACTGTAT	TGGAGAGTGC	AAGCGGAGAT TTAGTAGTAA	AGAACACTCT	ATTACTGCAG	AAATTTCTAT	TGGATAGCAC	TTATAATAGT	TTAGTGTAAC	TTAAAACTCC	219
TCARACTGGT	GACAATTTAA	CCCACAARCA	CCOLLOCOCO		WOOT I WORK!	CICIOCCUII	ACCOUNTAGE	WYZZZYYY	TTTGCAATTA	220 221
GAATGCCTGC	CATTTTTCAA	CATANTGGAT	GTAAGGTATT	ACACATATAC	CTGGGGATGG	GGAGGTAGGT	ATAATTGCAC	AAGCATTCTA	GGATTCTTGA	222
TCCCACACTC	TATCTTCCTT	THELCOCK	CRCTONCOLO			WINNEWIN	TTTGCTGGAC	CIGNICITAL	AACTCATAAA	223. 224
CATATCTTTA	CTCTTTTT	CALCECTORA	1 - 1		CCINIONOON	CIGCIACCCI	TACCINCOAT.	<b>VVCCCCLLCC</b>	TTCACTCATC	225.
ATGTATTTAA	TATCCATGTA	TCTATTCTCT	CTAATTTTGT	CATTITGTGT	TCTCATGTAT	TTTCATTCAT	TATGTGTCCA	ACTTCCATCG	TATGCCTTCA	226:
TACAACAAAA	GATACTTCC	TATGACAATT	ATCTTCCTTG	CCTTTCTCCC	ACATAGAACA	GTGCACAGAG	TAGGGGATCC	AAGAACCCAG	GAGAATATAT	227: 228:
										229:
75.01.0000			TVATCCCVCC	TALACUGGAG	GCTGAGGCAG	CAGAATCCCC	BC LLCCCCCC	1000001000		230:
AAGGATGTCT	GCTTTGTGAG	CCAGCCTGGG	CCACAGAGCG	AGACTCCGTC	TCAMMANA	********	amageccaag	AAAAATTTTA	AAAAAAAAA	2315
AAGGCTCAAT	TTAGTCACAT	TITAGCATIG CATTICCGTT CITATITICT	TCTCACCCAC	COCCTTTALA	TGAAATGGCA	AATACATTTA	AATCAGAACT	AAAAAGGGGA	ACAGGGTATA	232!
TTAGTAACTC	AAGCAGACAC	CTTATTTTCT	TTTCAAGCAG	AAAAGACTAT	GAGATGGTCG	TIGIGGITGE	TCCGGGAGGG	AGRAGATATA	GANAGNANGT	2335
GATCACTCAT	ACATGCATGT	GACCTCACTG	CACACTTATA	GTTATTGTAC	CTGTTGTCTT	TTTGCTGTCA	AGCCTAGCTA	AGATCATTTG	GAATGTTCAA	234! 2355
ACTITTATTC	TTTTCCAAAG	GCALCARCOT	CLCCCLCCCC	WIGHT CUCTE	CCINITICAT	CCACATGAAC	TANGATTACT	GATGTGTACA	GATTCAAAGC	2365
TCCAGTTCCT	TATGAATGGT	TACTGGTTTT GGATGGGCCA	CAAAAATATG	AGATAAATTG	AGTGTATAAA	AGTCATTTT	AGACAAAATG	AAACAGGAAA	ACCACCTCCA	2375
CCTCCTGTGA	CTANGGCATC	GGATGGGCCA AAGAGAAAGC	GCTCCACCAT	GTCATGGTTA	ATCTGCAGGG	AGGAAATACT	AGATTTGATT	GCAGATCAGA	CTGCAGCAAA	2385 23 <b>9</b> 5
CAGTGTCACC	TAGAAAAGAG	TOTTTCAAAA	TO CONTACTOR	2000000110	MATAGLANCE	ACAITATATA	TCTAGCTTTG	AAATATGAAA	TACTGTTTAG	2405
AAATAAAGTG	ATCACTTGGT	GAAGAAATCT	CACAAAGAAG	AACATAGAGA	GTTCACTTTC	ATCTGGAGTA	ATGARCAGAT	TGAACAAACT	CCAGGGTGGG	2415
GCAAGACGAT	TCTGACCTCC	TAGGTGAGCT	GTTTGCAAGA	GCCACAAGGG	AAAGGGGAAG	ACAACTICTT	TGTGGACTTA	AGGGTGAAAG	TTGCAAGCAG	2425 2435
CARATGITTG	TCGGAATTGT	TGACTTARAC	1 CCRC Record	CCURCUNCEN	CIGGOTICGT	TACGUAGGTT	GGGCAGCATT	GGGAGCAAAT	GTTGATTGAA	2445
CCAGAGATCA	GAGCAGGCTA	AGGGACTGCT	GGGATCCTGT	CCAGCTTTGA	GACCCTACAG	AGCCATGTTC	ACCTAGCACG	TATCCCCTCT	GCCCCCCCCC	2455
CCCCCTATCG	TTCCGGAACG	GGGCTTTCAC	CTCAGCTTGC	CAGGCTGGAG	CCAAGGGCCA	ACGCAGCCGC	CCCTTGTTCG	CGATGGTAGC	TTCCCAGGAG	2465: 2475;
GTTAGAAGGT	TCCGGACAGG	AACGGCGTGA	CCCCAATCCA	110CIACCIC	CIVVVCCCVV	AGGCACTGGC	GGGCCGGCC	AGCTTCTAAA	GTCGCCCAAG	2485
CGGGAAAGGA	AGCAGGGTCT	CTGAAGAAAT GTGTAGTAAA	ACTTCAGGAG	TAGAAAGAGG	AAGCTAGAGG	GTTAAATGCA	CTACACAGGA	ACAGARATEA	CTTGAGAACT	2495;
AGGGAAACTG	CAACGCCTGT	GTGTAGTAAA	CTAAAACAAG	TCTTGAATTG	CATACCGCCA	CGTAGGGAAG	AMATGAMAAC	CTTTGAATAT	TAGTGAAAA	25057 25157
AATTTGGTTT	GGATCCCATG	CCCATGACCC	TOCCARCTO	CALIBORNA	CARCCOACAGA	TTTAAAGAAG	CAACACCGCA	TITIGGCTTT	CTAAAGCTTT	25257
TGGCCCTTTA	TGTGAAGTAC	CTGGTTTTTC	CATTTTCTGT	TTTACCATAG	GCCTCAGTTC	GGTGTGTGGC	CTATTTATTC	CACATTTCCG	AAGAACTATT	25357
AACCGTATTA	AGATTAAAAA	AAAAGAATAC AATGTCCAGG	AATGGAAGCC	AAGTGATTAA	GCTTTCCTTA	TOCTTATATT	AAGTTGTAGC	ATATGCATTT	ACCGATAGTT	25451 25557
TCTCCATCCA	CTTCCCTCAG	CTTTGGCCTC	AAGCTATCTT	TALLCCTICA	CICITITION	CCTARAGRAA	AICTITAAAA	TGTCTTAGCA	TTTTCCCCAG	25657
TGTTCGATTA	GGACACATCT	CACTEGEAGA	TAACAGCCAA	I COMMITTEE	CIGIACAGC	1011000010	TACAGCTAGC	TACAGAGATT	CAATCCTTTC	25757
TAATCGCACC	TATAGTAATA	TCTTCACTTC	AAAAAGCCCT	CTATTATTCC	TATCTCAGAT	GATAAAAATT	CAATTAAGAG	AAATAAGAAC	GTGACATGTG	25857 25957
AAAAACCAAA	GTGAGCATCC	CATCTGTTCC	CACTCALLTC	TOOM CARL	VICCICIONO	ATTATTEME	CAAATGCATT	TCAATGACTA	GTTAACCATT	26057
CAAATGAATT	TECTTTETAT	ATGAGTGAGA	GCAAACACTC	TTTATTGTAC	AACTTGGGTG	GGTAAGTAGG	GAGAATAATG	GCTTTACTGA	ACATCACAAA	26157 26257
TGTATATTT	TIGGAGITAA	AGGTTAGGAA	GAAAACCAAA	GGGTAAGAGC	TGTTGTTCTG	GGCTGGCATT	GTCAATGAAG	AGCATAAATT	CAGATGTGAA	26357
TTGACAGATT	ATAACTCAGA	TGTCTTACTC	AGAGCATATG	CCTTCCCATT	TCTGAAAGAG	GGAAAACAGG	CTCCCATTAG	ACTATGACTA	ACAAAAATGT	26457
										26557 26657
CACCAGAAAG	TAGTAGAACG	TGGCAAGGAT	GTTTGGTCAG	GGGTTGGCAA	AAATAATGCT	CTTCAGACTT	AAAAGAACAC	AACCATATTT	CTTAGCCATC	26757
CCTGGCCTAA	CTAGCCTACT	GAGCTGAGAG	TOTAL CALLED	TCAGGAGTCA	GACTAGCTAC	ATCATAATCT	CICIGCCCVC	GGGCTGTGGA	TGTCATCCAT	26857
TTCCAATTGC	TTANACAAAT	ATGTTCAGTT	GTAACTATCA	ATACCACTAT	ATAACAGTGT	TGGCCAAGTT	TTATTGATGC	TGACAATCAA	TTGGAGTTAC	26957 27057
CTGAAGTCTG	ATGAGACAGC	CAGAGCATGT	CCTACCTACC	CICTROCACT	CORCAGGIC	GTGCATCTGA	GGTTTACTGC	TTTGCATTTT	TGTTTTGTAA	27157
										27257
										27357 27457
		CTTGGCCATG ACACATCCTG								27557
										27657
										27757 27857
CTGAGCCCCA	CANAGRATGE	ACCCAPTACE	CCCCCCCCCCC	TCTTCCATGG	CACITITCIT	CCTAGCTTTT	GAACAAGGGG	CCCCACATT	TTATTTCTCA	27957
										28057
										28157 28257
TTAATATCAT	GTCATAAGGA	TATTATGTTG	TATTALATGE	CTTTAAAACA	CCACAATGAT	TACCTCTAAA	CTTTAGCTCA	TGCATCTATA	ATAAATATA	28357
										28657 28557
	-4611 40CC00	GIGCHGICGE	CACGCCTGTA	ATCCCAGCAC	TTTGGGAGGC	CAAGGCGGGC	GGATCACGAG	CTCCACACAT	CCLCACCARC	28637
CIGOCCAACA	TEGTERARCE	CCATCTCTAC	TAMMATACA	AAAATTAACT	GGGCATGGTG	GCATGCGCCT	GTAGTCCCAG	GAGAATTCCT	TCAACCTCCC	28737
VOOCGONGC L	TOCAGTOAGE	CAAGATCTCA	CCACTGCTCT	CCAGCCTGGT	GACAGGGCAA	GACTCCGTCA	ALALABALA	AGAGAGGGAG	ACCCACACTA	28857
*********	AGTUAGAGCC	CTTTAATGAG	TO ACCUPACE	ACCTOTOCAC	COLCCIO	1000000010				28957
AGACTGTGTG	CACTITAATA	CAAGGGCAGT	CGTTCAGAAC	TACTCACCTC	CCCLLLACCE	ACTACCTCGT	TICCCITGII	TATGAATGGG	TTCATGCCTA	29057
										29157 29257
										29357
AATGAGCCCT	GCTATTCCTC	ACTGCCTGGA	TGGCTATAAG	CACAGCCCTT	ATTCACCCC	GAGGGGAGTA	GAATATGGTT	AAGAGAGAGT	GGAAAGAATG	29457 29557
										29657
										29757
		GAAGCCATTC AAAACACTAA								29157 29157
		***				WATERWAY I.E.		196	TATTITUE	4300
		TATTTATTTT						Val Val I		30053
Lys Val As	P Ala Phe C	ye Gly Gly i	Rer fla Usi	Ace Clare				***		30134
Lys Ile Th	r Val Val A					_				10223
		-								

Figure 8E

STEMFAM ATTETTETTE ANTALATTEE SCTALAGECA GALGEGTCAT ANTITEAGAL CECACETEES ACCEPTEED AGEATECAT MINISCECET ATTATEACTE ATTACATES GENELATA GETETTEATH TAGGELTITE CATACLAGAL GECETTECEA AMATEAGTG GICCITITIA TETETGETEE TIGGECALAE CIGTAGEAG TECTEAGALA ACALACATIT GARTALTEG CELAATGAGT TIGTGETELA AGEATACTE AMATITEGIA ANTITAGGAT ANTICATGAC TAGTGGATTE ATTACACCA ATGALGAGGET TATACAGGA TAGGTGAACA MICATAGTE CIGANIGGET THITGGICTG ALAMATATGE ATTGGCTCTC ATTACATTA ACCALLATA TECACATATA AGAATGAGAT	TCATGTCACC AAAAGGGGTG GAACCATCEC CTITAACATT 234	30329 30429 30529 30629 30729
CONTINGE TONORGEOU CANGENGION CITAGAAAAT CIGIGIATGE GAAAIACIGE TEGIGACITA ANATGAAAT TATTITTAAT	•	30826
BIS AST ILE GIU GIU THE GIU HIS THE GIU GIN LYS AEG AST VOL ILE AEG ILE ILE PEO HIS HIS AST TYF AS CAP ANT ATT GAG GAG ACA GAA CAT ACA GAG CAA AAG CGA AAT GTG ATT CGA ATT ATT CCT CAC CAC AAC TAC AA	T GCA GCT	30907
Ile Asn Lys Tyr Asn Bis Asp Ile Ala Leu Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pr NT AAT AAG TAC AAC CAT GAC ATT GCC CIT CTG GAA CTG GAC GAA CCC TTA GTG CTA AAC AGC TAC GTT ACA CC	T ATT TGC	30988
HE Als ASP Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr Val Ser Gly Trp Gly Arg Val Ph HT GCT GAC AAG GAA TAC ACG AAC ATC TTC CTC AAA TTT GGA TCT GGC TAT GTA AGT GGC TGG GGA AGA GTC TT	e Bis Lye C CAC AAA	31069
Cly Arg Ser Ala Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg Ser Thr Ly GG AGA TCA GCT TTA GTT CTT CAG TAC CTT AGA GTT CCA CTT GTT GAC CGA GCC ACA TGT CTT CGA TCT ACA AA		31150
HE TYP ABN ABN Net Phe Cys Als Gly Phe His Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pr MC TAT AAC AAC ATG TTC TGT GCT GGC TTC CAT GAA GGA GGT AGA GAT TCA TGT CAA GGA GAT AGT GGG GGA CC		31231
THE GIU VAI GIU GIY THE SEE PAG LEU THE GIY IIG IIG SEE TEP GIY GIU GIU CYS ALS MET LYS GIY LYS TY MET GAA GGG GAA GGG ACC AGT ITC TTA ACT GGA ATT ATT AGC TGG GGT GAA GAG TGT GCA ATG AAA GGC AAA TA		31312
THE LYS Val Ser Arg Tyr Val Asn Trp Ile Lys Glu Lys The Lys Leu The STOP THE ACC ANG GEN TOO COG THE GEC AND TOG ATT ANG GAN AND ACA ANG CTC ACT TAN TGANAGATGG ATTTCCANGG	TTAATTCATT	31399
CONTIGAR NITACAGGG COTOTACTA ACTANICACI TICCCATOTI TIGITAGATI TGANTATATA CATICTATGA TCATIGCTITACCOCCAGAAT TICATATITI ACCIGACAA ATTGATTAGA AAATGGAACC ACTAGAGGAA TATAATGTGT TAGGAAATTA CAGTCATTTC	TAAGGGCCCA	31499 31599
CONTIGUES ANATYGIGAN GITANATTOT CONCECTATO CATCAGATAC TATGGITCTC CACTATGGCA ACTARCTCAC TOATTITTCC CANTOCATC TECCCOATCT TOTTTCCTTC TOCANCOANA ACATCAATGT TIATTAGTTC TGTATACAGT ACAGGATCTT TGGTCTACTC	CTCCTTAGCA TATCACAAGG	31699 31799
CIGITACTAC ACTUATGANG ANAGANCACA GGAGTAGCTG AGAGGCTAAA ACTUATACTC CTTTTCCTCT ACCCTATTC	TO A A TOP TO THE	31899
MITTITCEA MATCCCAATC COCAMATCAG TITTITCTCTT TCTTACTCCC TCTCTCCTT TTACCCTCCA TGGTCGTTAA AGGAGAGATG TCTGTTATA CITCTGTACA CAGTTATACA TGTCTATCAA ACCCAGACTT GCTTCCATAG TGGAGACTTG CTTTTCAGAA CATAGGGATG	GGGAGCATCA	31999 32099
CTGAAAAGT TTGGGGGAAA AGTTTCTTTC AGAGAGTTAA GTTATTTTAT ATATATATA TATATATA	ATATATAGTG	32199
INTEGRATA TECESTATUTE TAGACACACA CECATACACA CATATAATOG AAGCAATAAG CCATTCTAAG AGCTTGTATG GTTATGGAGG	TCTGACTAGG	32299
CHIGHTTCA CGAAGGCAAG ATTGGCATAT CATTGTAACT AAAAAAGCTG ACATTGACCC AGACATATTG TACTCTTTCT AAAAATAATA TACAGAAAG AAGAGAACCG TTCGTTTGCA ATCTACAGCT AGTAGAGACT TTGAGGAAGA ATTCAACAGT GTGTCTTCAG CAGTGTTCAG	ATAATAATGC	32399 · 32499
ENGITORAG TIGCCINGRC CAGROGRACH ARGIRICATO ICICCITIAR CIRGCRÍNCO COGRRAGIGA GRAGGGIGCA GORGGCICAR	AGGCATAAGT	32599
CANCELATE AGCENACIAN GITGICETIT TETEGITTEG IGITEACENT GGNACATITI GATTATAGIT ANTECTICIN TETEGANCE	TCTAGAGAGT	32699
PRIGACCAA CIGACGIAIG ITICCCITIG IGAAITAATA AACIGGIGIT CIGGIICATA CCITGGCITI TIGIGGATIC CATIGATORG	AATCAGTCAC	32799
CTOTATTE ATGATOCATG GGACTACTGA CAAAATCACT CTGACCCTGC CAAGCTGCTG CCTTCTCCTG CCCCAACCTC ACCCCAGCC	AGGCCTCACT	32899
CHOCKAGET COTTEAGETIC TETTAGECAA TAKATTETTIG TOTTOGCATA TAAGTATAAA TAAACATATT TETAAATTEC TEGGOGGG MOOTATAA TOCCAGCACT TOTGGAGGGC AAGGTGGGGG GATCACCTGA GGTTAGGAGT TECAGGCGAG COTGGCCAAC ATGGTGAAAC		32999
STANAMATAG AACAATTAGC TGGGCTTGGT AATGTGCACC TATAATCCCA GCTACTGGGG AGGCTGAGGC AGGAGAATCA CTTGAGGCTG		33099 33199
GTGCGGGAG GTTGCAGTGA GACAAGATCG CACCAGTGCA CTCCCCATCC TGGGTGACAG AGTGAGACTC TGTCTCAAAG AAAATAAATA		33299
THETTGAGG COTTTETTGT TAAATCATTE ATGGAGAGGE ATCCCAAACA CEACATTEAA CAAAACACTE TGAAAAATGT TTTCAAATGG		33399
MEXICAGATT TGATGCTCTG TTATCCAGTT TTCATATAAG GCTGTGTGAG CTGTGTCCCA GAGAGGACAG TGGTCTGAAT CCACCTGAGA	CAGAATTGGG	33499
MINACTANC TOTGROTATG GOCTTCAATA AGTCACTCTC CATTTGGGAA TITGATTTCT CCACTTGTAT AATGAGAGTA TITGACAGG	TGCTCTCCCA	33599
MICCOTICC ANTITICITÀ CICTOTGATT ICATCITITI ATTITIATIC CITCATCCAA CAAATACTCA AGGACTAATI CCTGTGTGCC GTATICATI AAATTGTAAT TCAGATTITA TATATATATA AATAATGTAT AATGGTATA AATTGCTITG TGAGTGCCTA CTACACTGCT	AAATACCAAC	33699 33799
TATCANTAN CITCITAGCT GANTCAGANT CENTCITTAT CCCAGAGTNG CANTTRCTCT TGCNTCGAGT NTCGTGAANG ANGGCCACNC	TTAAATAAGA	33899
AMATGOCTO GGGTTTAGGT TTTATGAAAA AATGAAAGGA AATTAGTTCT GCTTTTGTTG ACTAAAGGAA GGGAAGAGAG AAGAGACACT CCTCAGATT TAAGGAGGAG GCTAATTCAT GCATTAAACA CGTTACTTCA AATTTGAATG ACCAAAGGTC TGTAGCCTCA GCACTTCAAA	ATAATTGTCT	33999 34099
97MGACACT CTGGCCTTGT TTCCATAGAG ACCACCCCTT ACAAAGGCAC CAATGGGAAA CTGGCCTCAG GACTCCTGTT ATTGGTCTTY	TOTGTGGCAG	34199
MANAGEAGE TETTGGACCE ATANATETET GAGECACAGT TETTTTTGGE ATGGGETENA ANATGATTGA ATTENTENTE AGCENCETGT	GGCATATTGC	34299
GENETALAE ATGTGGGGCC TITANGCTCA CTANGAGCCA ATGTCTTCAG AGCCAGCCCT GGCTTGATTC TACCTAGGGC ATTTGCAGTT UICATTAGT GCTTTCAAAA TTACTGTAGA TACTTTGCCT AAATAGACTA AAACATGCTG CCGTCATATT GGAAGTGACA GATTAAAATA	GCCATATAAG	34399 34499
UNSTRANG ANASTOTECT ANTATANTEC ASTCATTTA ACTICCTOT TANGESTAN TOTTTANAT CONTINUAT ACTICCTOR TANGESTAND TOTTTANATO	TTATACTCAC	34599
ACCANCERAG TACTOTOCAR TITTCTCTGC CARGGRARAR AGRARAGGTG TTCTTCCTTR CTTRCCTGRA CCRARACAGA CCRGTTTRCI	\ AAATTGCCTA	34699
MINIMATIG CTANACANGT TOCGNATGCT TACAGTCIAN TCCAAGANTG TCAGAGCTGC AAGGGCCCTT AAACACCATC CAATCCACTC CAGITGANG AGATTGAGGG CAACATAAGG CCAGGCCCAA GATAACACAA TGACAGCCAG GACTAGAGCT CAAGTCTCCC ACCCTGCACT	: CACTCATTA	34799 34899
MINITICA ACTGGAGIAC ATTAXCTCTA CIGICTATAT TITTAGGGCA GCTGGGGGAT TCTGCATGGG TGGCAATCCT CTCAACAAC	CTGGGACTGA	34999
MACTGCCTG GAATTCTTAC TAACAATTCT CTAATTGACC AAAAGGTGAC GAAATCAAGG AGACCAATAA GGTAGCCTTG GAAAGCAAG	GTGGC	35094

# Figure 9A

1	gaattccgtg	gatgtgcttt	aaaaccacac	ctaacgtttg	agcacaagtc	tcacgaactg
61	gcgctacaac	ttcatcagaa	acgaagtctc	caaatctgtc	caacgcaaaa	acacaaagga
121	gtctaatgac	taagtcttcc	aaccacaact	gtctgctgcg	cccggaaaac	aagccggggc
181	tctggggacc	cggggctcag	gccgcctcgc	tccggcctag.	cccgccacc	ttagttgtgt
241	cateceegg	gcatgctgag	catccccccg	cggctccggc	acagacgccc	ggacctcagg
301	tctctgcctc	cgcgcggggg	cccggccctg	tggccggagg	gagcggccgg	atggagcgga
361	ggatgaaagg	cggatacttg	gaccagcgag	tgccctacac	cttctgcagc	aaatctcccg
421	gaaatgggag	cttgggcgaa	gcgctgatgg	tcccgcaggg	aaagctcatg	gacccgggct
481	ccctgccgcc	ttccgactca	gaagatctct	tccaggacct	cagtcacttc	caagagacgt
541	ggctcgcaga	agctcaggta	ccggacagtg	atgagcagtt	tgttcctgat	ttccattcag
601	aaaacttagc	tttccatagc	cccaccacca	ggatcaagaa	ggaaccccag	agtccccgca
661	cagaccccgc	cctgtcctgc	agcaggaagc	caccactccc	ctaccaccat	ggagagcagt
721	gcctttactc	cagacaaatc	gccatcaagt	cccccgctcc	cggtgcccct	ggacagtcgc
781	ccctgcagcc	cttttccagg	gcagaacagc	agcagagcct	cctgagagcc	tccagctctt
841	cccagtccca	ccctggccac	gggtaccttg	gtgagcacag	ctccgtcttc	cagcagcccg
901	tggacatgtg	ccactccttc	acatctcctc	agggagggg	ccgggaacct	ctcccagccc
961	cctatcaaca	ccaactgtcg	gagccctgcc	caccctaccc	ccagcagaac	ttcaagcagg
1021	agtaccatga	cccctgtac	gaacaggctg	gccagcccgc	ttcaagccag	ggtggggtca
1081	gtgggcacag	gtacccaggg	gcgggggtgg	tgatcaaaca	ggagcgcaca	gacttcgcct
1141	acgactcaga	tgtccctgga	tgtgcatcaa	tgtacctcca	cccagagggc	ttctctggac
1201	cctctccagg	tgatggagtg	atgggttatg	gctatgaaaa	atcccttcga	ccattcccag
1261	atgatgtctg	cattgtccct	aaaaaatttg	aaggagacat	caagcaggaa	gggattggag
1321	ctttccggga	ggggccaccc	taccagcgcc	ggggtgcctt	acaactgtgg	cagtttctgg
1381	tggccctgct	ggatgaccca	acaaatgctc	atttcattgc	ttggacaggc	cggggaatgg
1441	agtttaaact	aattgaacct	gaagaggttg	ccaggctctg	gggtatccag	aagaaccggc
1501	cagccatgaa	ttatgacaag	ctgagccgct	cgctgcgata	ctattatgag	aaaggcatca
1561	tgcagaaggt	ggctggcgaa	cgctacgtgt	acaagtttgt	gtgcgagccg	gaggccctgt
1621	tetetetgge	cttcccagat	aatcaacgtc	cagctctgaa	ggctgagttt	gaccggccag
1681	tcagtgagga	ggacacagtc	cctttgtccc	acttggatga	gagtcctgcc	tacctcccag
1741	aactcactgg	ccccgctccg	cccttcggcc	acagaggtgg	atattcttac	taggcaccag
1801	tggcttcccc	ttgacatggt	ggggttgctc	agtgtatata	tcaactgatt	tggtattggt
1861	gaaggccctc	tttctgatgc	ctgtagaagt	ctctggggtc	agagctccac	tatcccatct
1921	gatactcctg	gccagactca	gctgctaacc	agagtctgcg	ggaaagacag	tggaggcagg
1981	ccaaatctaa	aggcagtagc	tgaagttcgc	tgtggctcac	ctgtaccttc	agttcagctt
2041	ggcctctgcc	taggtettge	tcagaggcca	agttcctcac	ccccaccaca	gagatccagt
2101	gttctattct	ggggacatac	agggacttcc	cttgtttatt	atggcaacag	ggccaagggg
2161	atteteagaa	caccctgtgt	crecetete	ccaaccccc	atgggagaca	aagttctgcc
2221	tggcttctgc	cctgaacagg	ggggtcctgt	gttcttggtg	ctgtgctctg	ggaggcagga
2281	gcatgtgggc	ggcagctggg	aaaaaarata	gaagtagaga	tggctctctg	ccctaggcct
2341	acccaggcct	aattccacct	ttgcctctta	tgccagacct	taataaagcc	tctgcttctc
2401	cccggaattc					

### Figure 9B

MTKSSNHNCLLRPENKPGLWGPGAQAASLRPSPATLVVSSPGHA
EHPPAAPAQTPGPQVSASARGPGPVAGGSGRMERRMKGGYLDQRVPYTFCSKSPGNGS
LGEALMVPQGKLMDPGSLPPSDSEDLFQDLSHFQETWLAEAQVPDSDEQFVPDFHSEN
LAFHSPTTRIKKEPQSPRTDPALSCSRKPPLPYHHGEQCLYSRQIAIKSPAPGAPGQS
PLQPFSRAEQQQSLLRASSSSQSHPGHGYLGEHSSVFQQPVDMCHSFTSPQGGGREPL
PAPYQHQLSEPCPPYPQQNFKQEYHDPLYEQAGQPASSQGGVSGHRYPGAGVVIKQER
TDFAYDSDVPGCASMYLHPEGFSGPSPGDGVMGYGYEKSLRPFPDDVCIVPKKFEGDI
KQEGIGAFREGPPYQRRGALQLWQFLVALLDDPTNAHFIAWTGRGMEFKLIEPEEVAR
LWGIQKNRPAMNYDKLSRSLRYYYEKGIMQKVAGERYVYKFVCEPEALFSLAFPDNQR
PALKAEFDRPVSEEDTVPLSHLDESPAYLPELTGPAPPFGHRGGYSY

# Figure 10 (A)

1	gaattccagg	ttggaggggc	ggcaacctcc	tgccagcctt	caggccactc	tectataeet
61	gccagaagag	acagagcttg	aggagagett	gaggagagca	ggaaaggtgg	aacattocto
121	ctactactca	ctcagttcca	caggtgggag	daacadcadd,	acttagagta	aggataetta
181	tacagataga	aaaacaaagg	cccadadadd	ggaagaaatg	cctaggages	222200000
241	adcasectes	accacagece	actactacaa	ctatazataa	atatagagee	accgagggca
301	cattcaccca	actacagece	agegeeggag	acatassass	acguagagua	geggaatate
201	cacccagcca	gctcagggga	aggacagggg	ecctgaagec	aggggatgga	gctgcaggga
201	agggagetea	gagagaaggg	gaggggagtc	tgageteagt	ttcccgctgc	ctgaaaggag
421	ggtggtacct	actcccttca	cagggtaact	gaatgagaga	ctgcctggag	gaaagctctt
481	caagtgtggc	ccaccccacc	ccagtgacac	cagcccctga	cacgggggag	ggagggcagc
541	atcaggaggg	gctttctggg	cacacccagt	acccgtctct	gagctttcct	tgaactgttg
601	cattttaatc	ctcacagcag	ctcaacaagg	tacataccgt	caccatcccc	attttacaga
661	tagggaaatt	gaggctcgga	gcggttaaac	aactcacctg	aggcctcaca	gccagtaagt
721	gggttccctg	gtctgaatgt	gtgtgctgga	ggatcctgtg	ggtcactcgc	ctggtagagc
781	cccaaggtgg	aggcataaat	gggactggtg	aatgacagaa	qqqqcaaaaa	tgcactcatc
841	cattcactct	gcaagtatct	acqqcacqta	caccaactcc	caagcaggtt	tacagattac
901	acaqcqqaqc	gatgcaatct	gatttagget	tttaaaggat	tgcaatcaag	tgggacccac
961	tagcctcaac	cctgtacctc	coctcoctc	cacccccage	agteteeaaa	agactacaaa
1021	aaccccagag	tgggggccat	gtatccaaag	aaactccaac	ctatatacaa	atcacactec
1081	ttttccagga	gcaaaaacag	322C2CCCtC	addetectag	cogtatacgg	testestest
1141	ctcaccagga	taggggggg	adacageeeg	aggetggtea	adattgaacc	ceeteetget
1201	teteteese	tagggggcag	accaagcaga	gggetgtgta	gacccacata	aagagcetae
1201	abattta	gcacttcacc	cgaggcactt	cacaagcacg	cccgggaatg	aaacttccaa
1201	cccccggga	tgcaggtgaa	acagttcctg	gttcagagag	grgaagegge	ctgcctgagg
1321	cagcacaget	cttctttaca	gatgtgcttc	cccacctcta	ccctgtctca	cggcccccca
1381	tgccagcctg	acggttgtgt	ctgcctcagt	catgctccat	ttttccatcg	ggaccatcaa
1441	gagggtgttt	gtgtctaagg	ctgactgggt	aactttggat	gagcggtctc	tccgctccga
1501	gcctgtttcc	tcatctgtca	aacgggctct	aacccactct	gatctcccag	ggcggcagta
1561	agtcttcagc	atcaggcatt	ttggggtgac	tcagtaaatg	gtagatcttg	ctaccagtgg
1621	aacagccact	aaggattctg	cagtgagagc	agagggccag	ctaagtggta	ctctcccaga
1681	gactgtctga	ctcacgccac	ccctccacc	ttggacacag	gacgctgtgg	tttctgagcc
1741	aggtacaatg	actcctttcg	gtaagtgcag	tggaagctgt	acactqccca	qqcaaaqcqt
1801	ccgggcagcg	taggcgggcg	actcagatcc	cagccagtgg	acttagecee	tatttactcc
1861	tccgataact	ggggtgacct	tggttaatat	tcaccagcag	cctcccccqt	tacccctcta
1921	gatccactgc	ttaaatacgg	acgaggacag	ggccctgtct	cctcaqcttc	aggcaccacc
1981	actgacctgg	gacagtgaat	cqtaaqtatq	cctttcactq	cgaggggttc	tggagaggct
2041	tccgagetcc	ccatggccca	ggcaggcagc	aggtetgggg	cadaadaaa	attatagaat
2101	gggtatccgc	ctgctgaggt	gcagggcaga	tagagagget	acaactaaac	tcctattttc
2161	ataataacag	cagccatgag	gattatatac	tatttcccaa	tectaceca	teccected
2221	gtacctcctg	gtggatacac	taatteetat	aagcagaagt	agetgeagg	atctaaatat
2281	gcagtcctgg	caccccagga	tagagaacac	canccaanat	3246343336	geeeaggeee
2341	caccatttc	tttctgtttg	cacageteet	ctatatata	acagcaacag	tatattatat
2401	cctataaccc	tcaccacctc	tectactect	tegeetgeeg	ggggcccccg	attatata
2461	taagaagatt	agattaga	coccactge	cygycatyca		cttctataga
2521	teesestate	aaggttcaga	9499991999	gaggaacgcc	ggeteacatt	etecatecee
2521	cccagacacg	accaggaaca	gacctgtgtc	ageeteagee	ccacaccaaa	argggeetee
2501	ccatgtaccg	tggacctctg	ggeceteetg	ceceagegga	ggacaggaag	ctgtgagggg
2041	tactgtcacc	cagggctcaa	getggeatte	ctgaataatc	gctctgcacc	aggccacggc
2701	Laageteagt	gegegattaa	gcctcataac	cctccaaggc	agttactagt	gtgattccca
2/61	ttttacagat	gaggaagatg	gggacagaga	ggtgaataac	tggccccaaa	tcacacacca
2821	. tccataattc	gggctcaggc	acctggctcc	agtccccaaa	ctcttgaacc	tggccctagt
2881	. gtcactgttt	ctcttgggtc	tcaggcgctg	gatggggaac	aggaaacctg	ggctgaactt
2941	. gaggcctctc	: tgatgctcqq	tqacttcaqa	caqttqctca	acctctctgt	tetettagge
3001	. aaaacatgat	aacctttqac	ttctqtccc	tcccctcacc	ccacccgacc	ttgatctctg
3061	. aagtgttgga	aggatttaat	ttttcctqca	ctgagttttg	gagacaggto	aaaaagatga
3121	. ccaaggccaa	ggtggccagt	ttcctataga	acgeetetaa	aagacctgca	gcaatagcag
3181	l caagaactgg	, tattctcgaq	aacttqctqc	gcagcaggca	cttcttggca	ttttatgtgt
3241	l atttaatttc	: acaatagctc	tatgacaaac	tccacctttc	tcatctccac	gaaactgagg
3301	l ttcagagagg	, ttaagtaact	tqtccaaqqt	cacacageta	atagcaagtt	gacgtggagc
3361	l aatctggcct	: cagageettt	aattttagco	acagactgat	getecetet	tcatttagcc
3421	l aggctgccto	tgaagttttc	: tgattcaaqa	cttctqqctt	cagetttota	cacagagatg
	<del>-</del>		5 .5	- 33	, ,	

### Figure 10 (B)

3481 attcaatgtc aggttttgga gcgaaatctg tttaatccca gacaaaacat ttaggattac 3541 atctcagttt tgtaagcaag tagctctgtg atttttagtg agttatttaa tgctctttgg 3601 ggctcaattt ttctatctat aaaatagggc taataatttg caccttatag ggtaagcttt 3661 gaggacagat tagatgatac ggtgcctgta aaacaccagg tgttagtaag tgtggcaatg 3721 atggtgacgc tgaggctgtg tttgcttagc atagggttag gcagctggca ggcagtaaac 3781 agttggataa tttaatggaa aatttgccaa actcagatgc tgttcactgc tgagcaggag 3841 ccccttcctg ctgaaatggt cctggggagt gcagcagget ctccgggaag aaatctacca 3901 teteteggge aggageteaa eetgtgtgea ggtacaggga gggetteete aeetggtgee 3961 cactcatgca tracgtcagt tattcctcat coctgtccaa aggattcttt tctccattgt 4021 acagetatga agetagtget caaagaagtg aagteattta eeceaggeee eetgeeagta 4081 agtgacaggg cetggtcaca ettgggttta tttattgece agtteaacag gttgtttgac 4141 cataggegag attetettee etgeaceetg eegggttget ettggteeet tattttatge 4201 teetggtag aaatggtgeg agattaggea gggagtggae getteeetgt eeetggeee 4261 gcaaagagtg ctcccacctg ccccgatccc agaaatgtca ccatgaagcc ttcattcttt 4321 tggtttaaag cttggcctca gtgtccgtac accatggggt ccttggccag atggcgactt 4381 tetectetee agtegeeete eeaggeaeta gettttagga gtgeagggtg etgeetetga 4441 tagaagggcc aggagagagc aggttttgga gacctgatgt tataaggaac agcttgggag 4501 gcataatgaa cccaacatga tgcttgagac caatgtcaca gcccaattct gacattcatc 4561 atctgagatc tgaggacaca gctgtctcag ttcatgatct gagtgctggg aaagccaaga 4621 cttgttccag ctttgtcact gacttgctgt atagcctcaa caaggccctg accetctetg 4681 ggcttcaaac tettcaetgt gaaaggagga aaccagagta ggtgatgtga caccaggaaa 4741 gatggatggg tgtgggggaa tgtgctcctc ccagctgtca cccctcgcc accctccctg 4801 caccageete tecaceteet tigageecag aatteceetg tetaggaggg cacetgtete 4861 gtgcctagcc atgggaattc tccatctgtt ttgctacatt gaacccagat gccattctaa 4921 ccaagaatce tggetgggtg caggggetet egeetgtaae eecageaett tgggaggeea 4981 aggcaggcgg atcaagaggt caggagttca agacctgcct ggccaacacg gtgaaacctc 5041 agctctacta aaaatacaaa aattagccag gcgtggtggc acacgcctgt aatcccagct 5101 atttgggaag ctgagacaga agaatttctt gaacccggga ggtggaggtt tcagtgagcc 5161 gagatcacgc cactgcactc caccetggcg gataaagcga gactetgtet caaaaaaaaac 5221 ccaaaaacct atgttagtgt acagagggc ccagtgaagt etteteccag ecceatttg 5281 cacaactggg gagagtgagg ecceaggace agaggattet tgetaaagge caagtggata 5341 gtgatggeec tgecaggeta gaagecaaaa eetetggeec tgagggeaatt 5401 agtgtcccca ccctgcagag gcccaactcc ctcctgacca ctgagccctg taatgatggg 5461 ggaattteca taageeatga aggaetgeae aaagtteagt tgggagtgaa agagaaatta 5521 aagggagatg gaaatataca gcactaattt tagcaccgtc ttcagttcta acaacactag 5581 ctagctgaag aaaatacaaa catgtattat gtaatgtgtg gtctgttcca tttggattac 5641 ttagaggcac gagggccaag gagaaaggtg gtggagagaa accagctttg cacttcattt 5701 gttgctttat tggaaggaaa cttttaaaag tccaaggggg ttgaagaatc tcaatatttg 5761 ttatttccag cttttttct ccagtttttc atttcccaaa ttcaaggaca cctttttctt 5821 tgtattttgt taagatgatg gttttggttt tgtgactagt agttaacaat gtggctgccg 5881 ggcatattet ceteagetag gaceteagtt třečeateřg tgaagaegge aggiteřace 5941 tagggggctg caggcaggtg gtccgaagcc tgggcatatc tggagtagaa ggatcactgt 6001 ggggcagggc aggttctgtg ttgctgtgga tgacgttgac tttgaccatt gctcggcaga 6061 gcctgctctc gctggttcag ccacaggccc caccactccc tattgtctca gccccgggta 6121 tgaaacatgt attectcact ggcctatcac ctgaagcctt tgaatttgca acacctgcca 6181 acceptect caaaagagtt geceteteta gateettttg atgtaaggtt tggtgttgag 6241 acttatttca ctaaattctc atacataaac atcactttat gtatgaggca aaatgaggac 6301 cagggagatg aatgacttgt cetggeteat acacetggaa agtgacagag teagattaga 6361 teetaggtet atetgaagtt aaaagaggtg tetttteaet teecacetee teeatetaet 6421 ttaaagcagc acaaacccct gctttcaagg agagatgagc gtctctaaag cccctgacag 6481 caagageeea gaaetgggae accattagtg acceagaegg caggtaaget gaetgcagga 6541 gcatcagect attettgtgt etgggaccae agageattgt ggggacagee eegtetettg 6601 ggaaaaaaac cctaagggct gaggatcctt gtgagtgttg ggtgggaaca gctcccagga 6661 ggtttaatca cagcccctcc atgctctcta gctgttgcca ttgtgcaaga tgcatttccc 6721 ttctgtgcag cagtttccct ggccactaaa tagtgggatt agatagaagc cctccaaggg 6781 ctccagcttg acatgattct tgattctgat ctgacccgat tctgataatc gtgggcaggc 6841 ccattcctct tcttgtgcct cattttcttc ttttgtaaaa caatggctgt accatttgca 6901 tettagggte attgeagatg aaagtgttge tgteeagage etgggtgeag gacetagatg 6961 taggattetg gttetgetae tteeteagtg acattgaata getgaeetaa tetetetgge 7021 tttggtttct tcatctgtaa aagaaggata ttagcattag cacctcacgg gattgttaca

# Figure 10 (C)

7081	agaaagcaat	gaattaacac	atgtgagcac	ggagaacagt (	gcttggcata	tggtaagcac
7141	tacgtacatt	ttgctattct	tctgattctt	tcagtgttac	tgatgtcggc	aagtacttgg
		tttaataatc				
		tgccttgact				
7321	atcttctatc	tcgtggggca	tcctcctact	gacagaceta	tactacctaa	tecetatete
7381	cctgactgag	gatccccagg	gagatgctgc	ccagaagaca	gatacatccc	accatgatca
7441	ggatcaccca	accttcaaca	agatcacccc	caacctggct	gagttcgcct	tcagcctata
7501	ccaccaacta	gcacaccagt	ccaacaggac	caatatette	ttctccccag	tgagcatcgc
		gcaatgctct				
7621	gggcctgaat	ttcaacctca	cocagattcc	adaggetgae	atccatgacg	acttccaaa
7681	actectecat	accctcaacc	agccagacag	ccadctccad	ctgaccaccg	gccccagga
7741	attecteege	gagggcctga	agetagtees	taagtttttg	gaggatatta	aaaaattata
7801	ccactcaga	gccttcactg	tcaacttccc	dageceda	gaggatgtta	aaaagutgta
7861	casttacata	gagaagggta	ctcaaccccgg	aattataaat	ttaataaaa	aacayaccaa
7001	agacacactt	tttgctctgg	traattarat	cttctttaaa	cctaacctta	agettgacag
0043	tasttass	tcccatagaa	acaaycaaaa	toottttaa	aaccaccage	terretate
0101	nananana	tgcattatgg	tttaataataat	tactted	cgtggattet	teagetteet
9101	acacacaaac	actaaaatgt	tttccatcat	tgagtaattt	gaggaaataa	tagattaaac
8101	tgtcaaaact	actgacgctc	tgcagaactt	ttcagagcct	ttaatgtcct	tgtgtatact
8221	gtatatgtag	aatatataat	gcttagaact	atagaacaaa	ttgtaataca	ctgcataaag
8281	ggatagtttc	atggaacata	ctttacacga	ctctagtgtc	ccagaatcag	tatcagtttt
		agacctgggt				
		acctctttga				
8461	ttgctgtcag	gtttttttag	gattaaacaa	atgacaaaga	ctgtggggat	ggtgtgtggc
8521	atacagcagg	tgatggactc	ttctgtatct	caggctgcct	tcctgcccct	gaggggttaa
8581	aatgccaggg	tcctgggggc	cccagggcat	tctaagccag	ctcccactgt	cccaggaaaa
		gaggggaggt				
8701	cccttgctct	tgaggcaaag	gagggcagtg	gaggcaagcc	aggctgcagt	cagcacagct
8761	aaagtcctgg	ctctgctgtg	gccttagtgg	gggcccaggt	ccctctccag	ccccagtctc
8821	ctccttctgt	ccaatgagaa	agctgggatc	aggggtccct	gaggcccctg	tccactctgc
8881	atgcctcgat	ggtgaagctc	tgttggtatg	gcagagggga	ggctgctcag	gcatctgcat
8941	ttcccctgcc	aatctagagg	atgaggaaag	ctctcaggaa	tagtaagcag	aatgtttgcc
9001	ctggatgaat	aactgagctg	ccaattaaca	aggggcaggg	agcettagae	agaaggtacc
9061	aaatatgcct	gatgctccaa	cattttattt	gtaatatcca	agacaccctc	aaataaacat
9121	atgattccaa	taaaaatgca	caqccacqat	ggcatctctt	agcctgacat	cqccacqatq
9181	tagaaattct	gcatcttcct	ctaqttttqa	attatcccca	cacaatcttt	ttcggcagct
9241	tagatagtca	gtttcagcac	cttttacaga	tgatgaagct	gagectegag	ggatgtgtgt
9301	catcaaggg	gctcagggct	tctcagggag	ggactcatg	gtttcttatt	ctgctacact
9361	cttccaaacc	ttcactcacc	cctaataata	cccaccttcc	cctctctcca	ggcaaatggg
9421	agagaccctt	tgaagtcaag	gacaccgagg	aagaggactt	ccacqtqqac	caggtgacca
		gcctatgatg				
9541	tatccaacto	ggtgctgctg	atgaaatacc	toggcaatoc	caccaccatc	ttetteetee
9601	ctgatgaggg	gaaactacag	cacctootaa	atgaactcac	ccacgatate	atcaccaagt
9661	tectogaaaa	tgaagacaga	aggtgattcc	ccaacctgag	aataaccaaa	aeactacca
9721	cacctcttac	ccatgttggg	actgaggccc	: atcaggactg	accadadad	tgaggagggt
9781	gaaccccaca	tecetgggte	actoctacto	totataaact	tagetteea	, egugguggge
9841	ccactgagtt	caggcagcgc	catecatact	cestasaase	eggooccoag	aacgaggeea
9901	gatassaata	tegteeetge	gectatge	. ctacgaggag	aacagtaccc	. agggcgagga
9961	tecaetece	z eegeceeegg	gaacttccca	a coccagogog	gacactgttt	, cttttttaata
10021	. cccagcgccc	aaggcaggga	cagcagcacc	accacacycc	ttassaat	Caaaaayyaa
10021	. cagacgggc	tcctggcaaa	tatataaa	g gagtgtggag	cccaagggta	gaatgteett
10101	. 3333333ac3	g gggaagagco	. igigiggdaa	a gycccagada	agcaaggitt	. yyaartyyaa
10141	ttagecagge	atgttcgcac	aaggettge	g cttetetgte	acttatcg	rgctgttaga
10201	l ctgggtgtc	tgtagtaagt	. gatacttaa	a caugagedad	acattagtgt	atgtgtgtgc
10201	tannanta	atgcccatgo	cetgetgate	cagutegete	gracactg	aaaaccaaga
10321	Legadataca	a aaaggtgtc	ggttcataa	aggaatcgag	gctggaatt	ctctgttcca
T0383	L Egccagcace	ccccgaggt	cctgctcca	g gggttgagaa	agaacaaaga	a ggctgagagg
1044	gtaacggat	c agagagecea	gagccagct	g ccgctcacac	cagaccctg	tcagggtggc
1050	attgtctcc	c catggaaaa	cagagagga	g cactcagcct	ggtgtggtc	a ctcttctctt
1056	atccactaa	a cggttgtca	tgggcactg	c caccagece	gtgtttctct	t gggtgtaggg
1062	r ccctdddda	t gttacaggc	c gggggccag	g tgacccaaca	a ctacagggc	a agatgagaca

## Figure 10 (D)

10681	ggcttccagg	acacctagaa	tatcagagga	ggtggcattt	caagcttttg	tgattcattc
10741	gatgttaaca	ttctttgact	caatgtagaa	gagctaaaag	tagaacaaac	caaaqccqaq
10801	ttcccatctt	agtgtgggtg	gaggacacag	gagtaagtgg	cagaaataat	caqaaaagaa
10861	aacacttgca	ctgtggtggg	tcccagaaga	acaagaggaa	tgctgtgcca	toccttoaat
10921	ttcttttctg	cacgacaggt	ctgccagctt	acatttaccc	aaactgtcca	ttactggaac
10981	ctatgatctg	aagagcgtcc	tgggtcaact	gggcatcact	aaggtcttca	qcaatqqqqc
11041	tgacctctcc	ggggtcacag	aggaggcacc	cctgaagctc	tccaaggtga	gatcaccctg
11101	acgaccttgt	tgcaccatgg	tatctgtagg	gaagaatgtg	tgggggctgc	agcactgtcc
11161	tgaggctgag	gaaggggccg	agggaaacaa	atgaagaccc	aggctgagct	cctgaagatg
11221	cccgtgattc	actgacacgg	gacggtgggc	aaacagcaaa	gccaggcagg	gactactata
11281	cagctggcac	tttcggggcc	tcccttgagg	ttgtgtcact	gaccctgaat	ttcaactttq
11341	cccaagacct	tctagacatt	gggccttgat	ttatccatac	tgacacagaa	aggtttgggc
11401	taagttgttt	caaaggaatt	tctgactcct	tcgatctgtg	agatttggtg	tctqaattaa
11461	tgaatgattt	cagctaaagt	gacacttatt	ttggaaaact	aaaggcgacc	aatqaacaac
11521	ctgcagttcc	atgaatggct	gcattatctt	ggggtctggg	cactgtgaag	gtcactgcca
11581	gggtccgtgt	cctcaaggag	cttcaagccg	tgtactagaa	aggagagagc	cctqqaqqca
11641	gacgtggagt	gacgatgctc	ttccctgttc	tgagttgtgg	gtgcacctga	gcagggggag
11701	aggcgcttgt	caggaagatg	gacagagggg	agccagcccc	atcagccaaa	gccttgagga
11761	ggagcaaggc	ctatgtgaca	gggagggaga	ggatgtgcag	ggccagggcc	qtccaqqqq
11821	agtgagcgct	tcctgggagg	tgtccacgtg	agccttgctc	gaggcctggg	atcagcetta
11881	caacgtgtct	ctgcttctct	cccctccagg	ccgtgcataa	ggctgtgctg	accatcgacg
11941	agaaagggac	tgaagctgct	ggggccatgt	ttttagaggc.	catacccatg	tctatcccc
12001	ccgaggtcaa	gttcaacaaa	ccctttgtct	tcttaatgat	tgaacaaaat	accaaqtctc
12061	ccctcttcat	gggaaaagtg	gtgaatccca	cccaaaaata	actgcctctc	gctcctcaac
12121	ccctcccctc	catccctggc	cccctccctg	gatgacatta	aagaagggtt	gagctggtcc
12181	ctgcctgcat	gtgatctgta	aatccctggg	atgttttctc	tg	

## Figure 11

1	caccagcatc	atctcctcca	attcatccag	ctactctqcc	catgaagata	atagttttca
61	ggcggattgc	ctcagatcac	actatctcca	cttqcccaqc	cctqtqqaaq	attaggggg
121	atgtattcca	atgtgatagg	aactgtaacc	tctggaaaaa	qqaaqqttta	tettttatee
TRI	ttgctgctca	ttggcttctg	ggactgcgtg	acctqtcacq	ggagccctgt	ggacatetge
241	acagccaagc	cgcgggacat	tcccatgaat	cccatgtgca	tttaccqctc	cccggagaag
301	aaggcaactg	aggatgaggg	ctcagaacag	aagatcccqq	aggccaccaa	ccaacatatc
36 T	tgggaactgt	ccaaggccaa	ttcccgcttt	gctaccactt	tctatcagca	cctggcagat
421	tccaagaatg	acaatgataa	cattttcctg	tcacccctga	gtatctccac	gacttttact
481	atgaccaagc	tgggtgcctg	taatgacacc	ctccaqcaac	tgatggaggt	atttaagttt
541	gacaccatat	ctgagaaaac	atctgatcag	atccacttct	tctttqccaa	actgaactgc
POI	cgactctatc	gaaaagccaa	caaatcctcc	aagttagtat	caqccaatcq	cctttttgga
PPT	gacaaatccc	ttaccttcaa	tgagacctac	caggacatca	gtgagttggt	atatggagcc
721	aagctccagc	ccctggactt	caaggaaaat	gcagagcaat	ccagagegge	catcaacaaa
<b>78T</b>	tgggtgtcca	ataagaccga	aggccgaatc	accgatgtca	ttccctcqqa	agccatcaat
841	gageteactg	ttctggtgct	ggttaacacc	atttacttca	agggcctgtg	gaagtcaaag
901	ttcagccctg	agaacacaag	gaaggaactg	ttctacaagg	ctgatggaga	atcatattca
961	gcatctatga	tgtaccagga	aggcaagttc	cgttatcggc	gcgtggctga	aggcacccag
1021	gtgcttgagt	tgcccttcaa	aggtgatgac	atcaccatgg	tcctcatctt	gcccaagcct
TORT	gagaagagcc	tggccaaggt	ggagaaggaa	ctcaccccag	aggtgctgca	ggagtggctg
1141	gatgaattgg	aggagatgat	gctggtggtt	cacatgcccc	getteegeat	tgaggacggc
1201	ttcagtttga	aggagcagct	gcaagacatg	ggccttgtcg	atctqttcaq	ccctgaaaag
1261	tccaaactcc	caggtattgt	tgcagaaggc	cgagatgacc	tctatqtctc	agatgcattc
1321	cataaggcat	ttcttgaggt	aaatgaagaa	ggcagtgaag	cagctgcaag	taccactatt
TRRT	gtgattgctg	gccgttcgct	aaaccccaac	agggtgactt	tcaaqqccaa	caggeeette
1441	ctggttttta	taagagaagt	tcctctgaac	actattatct	tcatgggcag	agtagccaac
1501	ccttgtgtta	agtaaaatgt	tcttattctt	tgcacctctt	cctatttttg	gtttgtgaac
1561	agaagtaaaa	ataaatacaa	actacttcca	tctcacatt	_	

## Figure 12 A

1	ctgcaggggg	999999999	gggggctgtc	atggcggcag	gacggcgaac	ttgcagtatc
				ctccagaatg		
121	gctgttcgtg	gccacctggg	gaatttccgg	cacaccagct	cctcttgact	cagtgttctc
181	cagcagcgag	cgtgcccacc	aggtgctgcg	gatccgcaaa	cgtgccaact	ccttcctgga
241	ggagctccgt	cacagcagcc	tggagcggga	gtgcatagag	gagatctgtg	acttcgagga
301	ggccaaggaa	attttccaaa	atgtggatga	cacactggcc	ttctggtcca	agcacgtcga
				gcacccgtgc		
				cagctgcgac		
481	ccgcttctgc	cagcgcgagg	tgagcttcct	caattgctcg	ctggacaacg	gcggctgcac
541	gcattactgc	ctagaggagg	tgggctggcg	gcgctgtagc	tgtgcgcctg	gctacaagct
601	gggggacgac	ctcctgcagt	gtcaccccgc	agtgaagttc	ccttgtggga	ggccctggaa
661	gcggatggag	aagaagcgca	gtcacctgaa	acgagacaca	gaagaccaag	aagaccaagt
721	agatccgcgg	ctcattgatg	ggaagatgac	caggcgggga	gacagcccct	ggcaggtggt
781	cctgctggac	tcaaagaaga	agctggcctg	cggggcagtg	ctcatccacc	cctcctgggt
841	gctgacagcg	gcccactgca	tggatgagtc	caagaagctc	cttgtcaggc	ttggagagta
901	tgacctgcgg	cgctgggaga	agtgggagct	ggacctggac	atcaaggagg	tcttcgtcca
961	ccccaactac	agcaagagca	ccaccgacaa	tgacatcgca	ctgctgcacc	tggcccagcc
1021	cgccaccctc	tcgcagacca	tagtgcccat	ctgcctcccg	gacagcggcc	ttgcagagcg
1081	cgagctcaat	caggccggcc	aggagaccct	cgtgacgggc	tggggctacc	acagcagccg
1141	agagaaggag	gccaagagaa	accgcacctt	cgtcctcaac	ttcatcaaga	ttcccgtggt
1201	cccgcacaat	gagtgcagcg	aggtcatgag	caacatggtg	tctgagaaca	tgctgtgtgc
						ccatggtcgc
1321	ctccttccac	ggcacctggt	tcctggtggg	cctggtgagc	tggggtgagg	gctgtgggct
1381	ccttcacaac	tacggcgttt	acaccaaagt	cagccgctac	ctcgactgga	tccatgggca
						ccctgcaggg
						cacaccggcc
1561	tgctgttctg	tccttccatc	cctcttttgg	gctcttctgg	agggaagtaa	catttactga
						caactctgtg
						tgaggggat
						aaaaaaaaa
1801	aaaaaaaaa	aaaaaaaccc	ccccccgccc	ccccccctg	cag	

### Figure 12 B

1 agtgaatctg ggcgagtaac acaaaacttg agtgtcctta cctgaaaaat agaggttaga 61 gggatgetat gtgccattgt gtgtgtgtgt tgggggtggg gattgggggt gatttgtgag 121 caattggagg tgagggtgga gcccagtgcc cagcacctat gcactgggga cccaaaaagg 181 agcatcttct catgatttta tgtatcagaa attgggatgg catgtcattg ggacagcgtc 241 tittitctig tatggtggca cataaataca tgtgtcttat aattaatggt attitagatt 301 tgacgaaata tggaatatta cctgttgtgc tgatcttggg caaactataa tatctctggg 361 caaaaatgtc cccatctgaa aaacagggac aacgttcctc cctcagccag ccactatggg 421 getaaaatga gaccacatet gteaagggtt ttgeceteae eteceteeet getggatgge 481 atcettggta ggcagaggtg ggcttcgggc agaacaagcc gtgctgagct aggaccagga 541 gtgctagtgc cactgtttgt ctatggagag ggaggcctca gtgctgaggg ccaagcaaat 601 atttgtggtt atggattaac tcgaactcca ggctgtcatg gcggcaggac ggcgaacttg 661 cagtatetec acgacccgcc cctgtgagtc cccctccagg caggtctatg aggggtgtgg 721 agggaggget geeceeggga gaagagaget aggtggtgat gagggetgaa teeteeagee 781 agggtgetea acaageetga gettggggta aaaggacaca aggeeeteea caggeeagge 841 ctggcagcca cagtctcagg tccctttgcc atgcgcctcc ctctttccag gccaagggtc 901 cccaggccca gggccattcc aacagacagt ttggagccca ggaccttca ttctccccac 961 cccacttcca cctttggggg tgtcggattt gaacaaatct cagaagcggc ctcagaggga 1021 gtcggcaaga atggagagca gggtccggta gggtgtgcag aggccacgtg gcctatccac 1081 tggggagggt teettgatet etggeeacca gggetatete tgtggeettt tggageaacc 1141 tggtggtttg gggcaggggt tgaatttcca ggcctaaaac cacacaggcc tggccttgag 1201 tcctggctct gcgagtaatg catggatgta aacatggaga cccaggacct tgcctcagtc 1261 ttccgagtct ggtgcctgca gtgtactgat ggtgtgagac cctactcctg gaggatgggg 1321 gacagaatet gategateee etgggttggt gaetteeetg tgeaateaac ggagaceage 1381 aagggttgga tttttaataa accacttaac teeteegagt eteagtttee eeetetatga 1441 aatggggttg acagcattaa taactacctc ttgggtggtt gtgagcctta actgaagtca 1501 taatatetea tgittaetga geatgageta tgigeaaage etgittigag ageittaigt 1561 ggactaacte etttaattet cacaacacee tttaaggeae agatacacea egttatteea 1621 tccattttac aaatgaggaa actgaggcat ggagcagtta agcatcttgc ccaacattgc 1681 cctccagtaa gtgctggagc tggaatttgc accgtgcagt ctggcttcat ggcctgccct 1741 gtgaatcetg taaaaattgt ttgaaagaca ccatgagtgt ccaatcaacg ttagctaata 1801 ttctcagccc agtcatcaga ccggcagagg cagccacccc actgtcccca gggaggacac 1861 aaacatcctg gcaccctctc cactgcattc tggagctgct ttctaggcag gcagtgtgag 1921 ctcagccca cgtagagcgg gcagccgagg ccttctgagg ctatgtctct agcgaacaag 1981 gaccctcaat tccagcttcc gcctgacggc cagcacacag ggacagccct ttcattccgc 2041 ttccacctgg gggtgcaggc agagcagcag cgggggtagc actgcccgga gctcagaagt 2101 cctcctcaga caggtgccag tgcctccaga atgtggcagc tcacaagcct cctgctgttc 2161 gtggccacct ggggaatttc cggcacacca gctcctcttg gtaaggccac cccaccccta 2221 ccccgggacc cttgtggcct ctacaaggcc ctggtggcat ctgcccaggc cttcacagct 2281 tccaccatct ctctgagccc tgggtgaggt gaggggcaga tgggaatggc aggaatcaac 2341 tgacaagtee caggtaggee agetgeeaga gtgeeacaea ggggetgeea gggeaggeat 2401 gcgtgatggc agggagcccc gcgatgacct cctaaagctc cctcctccac acggggatgg 2461 teacagagte ceetgggeet teceteteca eccaeteact ceetcaactg tgaagaceee 2521 aggcccaggc taccgtccac actatccagc acagcctccc ctactcaaat gcacactggc 2581 ctcatggctg ccctgcccca accetttec tggtctccac agccaacggg aggaggccat 2641 gattettggg gaggteegea ggeacatggg cecetaaage cacaccagge tgttggttte 2701 atttgtget ttatagaget gtttatetge ttgggacetg cacetecace ettteecaag 2761 gtgccctcag ctcaggcata ccctcctcta ggatgccttt tcccccatcc cttcttgctc 2821 acaccccaa cttgatctct ccctcctaac tgtgccctgc accaagacag acacttcaca 2881 gageccagga cacacetggg gaccettect gggtgatagg tetgtetate etecaggtgt 2941 ccctgcccaa ggggagaagc atggggaata cttggttggg ggaggaaagg aagactgggg 3001 ggatgtgtca agatggggct gcatgtggtg tactggcaga agagtgagag gatttaactt 3061 ggcagcettt acagcagcag ccagggettg agtacttate tetgggecag getgtattgg 3121 atgttttaca tgacggtctc atccccatgt ttttggatga gtaaattgaa ccttagaaag 3181 gtaaagacac tggctcaagg tcacacagag atcggggtgg ggttcacagg gaggcetgtc 3241 catctcagag caaggettcg tcctccaact gccatctgct tcctggggag gaaaagagca 3301 gaggacccet gcgccaagcc atgacctaga attagaatga gtcttgaggg ggcggagaca 3361 agaccttccc aggctctccc agctctgctt cctcagaccc cctcatggcc ccagcccctc 3421 ttaggeceet caccaaggtg ageteceete cetecaaaae cagaeteagt gttetecage 3481 agcgagcgtg cccaccaggt gctgcggatc cgcaaacgtg ccaactcctt cctggaggag 3541 ctccgtcaca gcagcctgga gcgggagtgc atagaggaga tctgtgactt cgaggaggcc

# Figure 12 B (continued)

3601	aaggaaattt	tccaaaatgt	ggatgacaca	gtaaggccac	catgggtcca	gaggatgagg
3661	ctcaggggcg	agctggtaac	cagcaggggc	ctcgaggagc	aggtggggac	tcaatgctga
3721	ggccctctta	ggagttgtgg	gggtggctga	gtggagcgat	taggatgctg	gccctatgat
3781	gtcggccagg	cacatgtgac	tgcaagaaac	agaattcagg	aagaagctcc	aggaaagagt
3841	gtggggtgac	cctaggtggg	gactcccaca	gccacagtgt	aggtggttca	gtccaccctc
3901	cagccactgc	tgagcaccac	tgcctccccg	tcccacctca	caaagagggg	acctaaagac
3961	caccctgctt	ccacccatgc	ctctgctgat	cagggtgtgt	gtgtgaccga	aactcacttc
4021	tgtccacata	aaatcgctca	ctctgtgcct	cacatcaaag	ggagaaaatc	tgattgttca
4081	gggggtcgga	agacagggtc	tgtgtcctat	ttgtctaagg	gtcagagtcc	tttggagccc
4141	ccagagtcct	gtggacgtgg	ccctaggtag	tagggtgagc	ttggtaacgg	ggctggcttc
4201	ctgagacaag	gctcagaccc	gctctgtccc	tggggatcgc	ttcagccacc	aggacctgaa
4261	aattgtgcac	gcctgggccc	ccttccaagg	catccaggga	tgctttccag	tggaggcttt
4321	cagggcagga	gaccctctgg	cctgcaccct	ctcttgccct	cagcctccac	ctccttgact
4381	ggacccccat	ctggacctcc	atccccacca	cctctttccc	cagtggcctc	cctggcagac
4441	accacagtga	ctttctgcag	gcacatatct	gatcacatca	agtccccacc	gtgctcccac
4501	ctcacccatg	gtctctcagc	cccagcagcc	ttggctggcc	tctctgatgg	agcaggcatc
4561	aggcacaggc	cgtgggtctc	aacgtgggct	gggtggtcct	ggaccagcag	cagccgccgc
4621	agcagcaacc	ctggtacctg	gttaggaacg	cagaccctct	gcccccatcc	tcccaactct
4681	gaaaaacact	ggcttaggga	aaggcgcgat	gctcaggggt	ccccaaagc	ccgcaggcag
4741	agggagtgat	gggactggaa	ggaggccgag	tgacttggtg	agggattcgg	gtcccttgca
4801	tgcagaggct	gctgtgggag	cggacagtcg	cgagagcagc	actgcagctg	catggggaga
4861	gggtgttgct	ccagggacgt	gggatggagg	ctgggcgcgg	gcgggtggcg	ctggagggcg
4921	ggggagggc	agggagcacc	agctcctagc	agccaacgac	catcgggcgt	cgatccctgt
4981	ttgtctggaa	gccctcccct	cccctgcccg	ctcacccgct	gccctgcccc	acccgggcgc
5041	gcccctccgc	acaccggctg	caggagcctg	acgctgcccg	ctctctccgc	agctggcctt
5101	ctggtccaag	cacgtcggtg	agtgcgttct	agatccccgg	ctggactacc	ggcgcccgcg
5161	cccctcggga	tctctggccg	ctgaccccct	accccgcctt	gtgtcgcaga	cggtgaccag
5221	tgcttggtct	tgcccttgga	gcacccgtgc	gccagcctgt	gctgcgggca	cggcacgtgc
5281	atcgacggca	tcggcagctt	cagctgcgac	tgccgcagcg	gctgggaggg	ccgcttctgc
5341	cagcgcggtg	agggggagag	gtggatgctg	gegggeggeg	gggcggggct	ggggccgggt
5401	tgggggcgcg	gcaccagcac	cagctgcccg	cgccctcccc	tgcccgcaga	ggtgagcttc
5461	ctcaattgct	ctctggacaa	cggcggctgc	acgcattact	gcctagagga	ggtgggctgg
5521	cggcgctgta	gctgtgcgcc	tggctacaag	ctgggggacg	acctcctgca	gtgtcacccc
5581	gcaggtgaga	agcccccaat	acatcgccca	ggaatcacgc	tgggtgcggg	gtgggcaggc
5641	ccctgacggg	cgcggcgcgg	ggggctcagg	agggtttcta	gggagggagc	gaggaacaga
5701	gttgagcctt	ggggcagcgg	cagacgcgcc	caacaccggg	gccactgtta	gcgcaatcag
5761	cccgggagct	gggcgcgccc	tccgctttcc	ctgcttcctt	tetteetgge	gtccccgctt
5821	cctccgggcg	cccctgcgac	ctggggccac	ctcctggagc	gcaagcccag	tggtggctcc
5881	gctccccagt	ctgagcgtat	ctggggcgag	gcgtgcagcg	tcctcctcca	tgtagcctgg
5941	ctgcgttttt	ctctgacgtt	gtccggcgtg	catcgcattt	ccctctttac	ccccttgctt
6001	ccttgaggag	agaacagaat	cccgattctg	ccttcttcta	tattttcctt	tttatgcatt
6061	ttaatcaaat	ttatatatgt	atgaaacttt	aaaaatcaga	gttttacaac	tcttacactt
6121	tcagcatgct	gttccttggc	atgggtcctt	ttttcattca	ttttcataaa	aggtggaccc
6181	ttttaatgtg	, gaaattccta	tcttctgcct	ctagggcatt	tatcacttat	ttcttctaca
6241	atctcccctt	: tacttcctct	attttctctt	tctggacctc	ccattattca	gacctctttc
6301	ctctagtttt	attgtctctt	ctatttccca	tctctttgac	tttgtgtttt	ctttcaggga
6361	actttcttt	: ttttctttt	ttttgagatg	gagtttcact	cttgttgtcc	caggctggag
6421	. tgcaatgacg	, tgatctcago	tcaccacaac	: ctccgcctcc	tggattcaag	cgattctcct
6481	gccgcagcct	: cccgagtagc	: tgggattaca	ı ggcatgcgcc	accacqccca	gctaattttg
6541	. tgtttttagt	agagaaggg	tttctccgtg	, ttggtcaagc	tggtcttgaa	ctcctgacct
6601	. caggtgatco	: acctgccttg	geeteetaaa	gtgctgggat	tacaggcgtg	agccaccgcg
6661	. cccagcctct	: ttcagggaad	: tttctacaac	: tttataattc	aattcttctq	cagaaaaaaa
6721	. tttttggcca	a ggctcagtag	, ctcagaccaa	a taattccagc	actttgagag	gctgaggtgg
6781	gaggattgct	tgagettggg	g agtttgagad	tagectggge	aacacaqtqa	gaccetgtet
6841	. ctattttaa	a aaaaagtaaa	a aaaagatcta	a aaaatttaac	tttttattt	gaaataatta
6901	l gatatttcca	a ggaagctgca	aagaaatqc	tagtagacct	attaactata	qatttcctac
6961	L aaggccgtgg	g gaaggccctg	g tcattggcad	aaccccagat	cqtqaqqqct	ttccttttag
7021	l gctgctttc	t aagaggacto	ctccaagct	: ttggaggatq	gaagacgctc	acccatagta
7083	l ttcggcccc1	t cagagcagg	tggggcagg	g gagetggtge	ctgtgcaggc	tqtqqacatt
7141	l tgcatgacte	c cctgtggtca	a gctaagagc	a ccactcctto	ctgaagcggc	gcctgaagtc
		•				=

### Figure 12 B (continued)

7201 cctagtcaga gcctctggtt caccttctgc aggcagggag aggggagtca agtcagtgag 7261 gagggettte geagtttete ttacaaacte teaacatgee eteccacetg cactgeette 7321 ctggaagece cacagectee tatggtteeg tggteeagte etteagette tgggegeece 7381 catcacgggc tgagattttt gctttccagt ctgccaagtc agttactgtg tccatccatc 7441 tgctgtcagc ttctggaatt gttgctgttg tgccctttcc attcttttgt tatgatgcag 7501 ctcccctgct gacgacgtcc cattgctctt ttaagtctag atatctggac tgggcattca 7561 aggcccattt tgagcagagt cgggctgacc tttcagccct cagttctcca tggagtatgc 7621 getetetet tggcagggag geeteacaaa catgecatge etattgtage agetetecaa 7681 gaatgeteac eteettetee etgtaattee ttteetetgt gaggagetea geageateee 7741 attatgagac cttactaatc ccagggatca cccccaacag ccctggggta caatgagctt 7861 actcttgcca ttgggtggta ctgtttgttg actgactgac tgactgactg gagggggttt 7921 gtaattigta teteagggat tacceccaac agecetgggg tacaatgage etteaagaag 7981 tttaacaacc tatgtaagga cacacagcca gtgggtgatg ctgcctggtc tgactcttgc 8041 cattcagtgg cactgtttgt tgactgactg actgactgac tggctgactg gagggggttc 8101 atagetaata ttaatggagt ggtetaagta teattggtte ettgaaceet geaetgtgge 8161 aaagtggccc acaggctgga ggaggaccaa gacaggaggg cagtctcggg aggagtgcct 8221 ggcaggcccc tcaccacctc tgcctacctc agtgaagttc ccttgtggga ggccctggaa 8281 geggatggag aagaagegea gteacetgaa aegagacaca gaagaccaag aagaccaagt 8341 agateegegg eteattgatg ggaagatgae caggegggga gaeageeeet ggeaggtggg 8401 aggegaggea geaeeggete gteaegtget gggteeggga teaetgagte cateetggea 8461 gctatgctca gggtgcagaa accgagaggg aagcgctgcc attgcgtttg ggggatgatg 8521 aaggtggggg atgetteagg gaaagatgga egeaacetga ggggagagga geageeaggg 8581 tgggtgaggg gaggggcatg ggggcatgga ggggtctgca ggagggaggg ttacagtttc 8641 taaaaagage tggaaagaca etgetetget ggegggattt taggeagaag eeetgetgat 8701 gggagagggc taggagggag ggccgggcct gagtacccct ccagcctcca catgggaact 8761 gacacttact gggttcccct ctctgccagg catgggggag ataggaacca acaagtggga 8821 gtatttgccc tggggactca gactctgcaa gggtcaggac cccaaagacc cggcagccca 8881 gtgggaccac agccaggacg gcccttcaag ataggggctg agggaggcca aggggaacat 8941 ccaggcagec tgggggccac aaagtettee tggaagacae aaggeetgee aageetetaa 9001 ggatgagagg agctcgctgg gcgatgttgg tgtggctgag ggtgactgaa acagtatgaa 9061 cagtgcagga acagcatggg caaaggcagg aagacaccct gggacaggct gacactgtaa 9121 aatgggcaaa aatagaaaac gccagaaagg cctaagccta tgcccatatg accagggaac 9241 gtgatgtcat catcccaccc cattccaggt ggtcctgctg gactcaaaga agaagctggc 9301 ctgcggggca gtgctcatcc acccctcctg ggtgctgaca gcggcccact gcatggatga 9361 gtccaagaag ctccttgtca ggcttggtat gggctggagc caggcagaag ggggctgcca 9421 gaggcctggg tagggggacc aggcaggctg ttcaggtttg ggggaccccg ctccccaggt 9481 gettaageaa gaggettett gageteeaca gaaggtgttt ggggggaaga ggeetatgtg 9541 ccccacct gcccaccat gtacaccag tattttgcag tagggggttc tctggtgcc 9601 tcttcgaatc tgggcacagg tacctgcaca cacatgtttg tgaggggcta cacagacctt 9661 cacctctcca ctcccactca tgaggagcag gctgtgtggg cctcagcacc cttgggtgca 9721 gagaccagca aggcctggcc tcagggctgt gcctcccaca gactgacagg gatggagctg 9781 tacagaggga gccctagcat ctgccaaagc cacaagctgc ttccctagca ggctgggggc 9841 tectatgeat tggeceegat etatggeaat ttetggaggg ggggtetgge teaactettt 9901 atgecaaaaa gaaggeaaag catattgaga aaggecaaat teacatttee tacageataa 9961 tetatgecag tggccccgtg gggcttggct tagaattece aggtgctett cecagggaac 10021 catcagtetg gactgagagg acettetete teaggtggga eceggeeetg teeteeetgg 10081 cagtgccgtg ttctgggggt cetectetet gggteteact geecetgggg tetetecage 10141 tacetttget ceatgtteet ttgtggetet ggtetgtet tggggtttee aggggteteg 10201 ggcttccctg ctgcccattc cttctctggt ctcacggctc cgtgactcct gaaaaccaac 10261 cagcatecta eccetttgga ttgacacetg ttggecacte ettetggeag gaaaagteae 10321 cgttgatagg gttccacggc atagacaggt ggctccgcgc cagtgcctgg gacgtgtggg 10381 tgcacagtct ccgggtgaac cttcttcagg ccctctccca ggcctgcagg ggcacagcag 10441 tgggtgggcc tcaggaaagt gccactgggg agaggctccc cgcagcccac tctgactgtg 10501 ccctctgccc tgcaggagag tatgacctgc ggcgctggga gaagtgggag ctggacctgg 10561 acatcaagga ggtcttcgtc caccccaact acagcaagag caccaccgac aatgacatcg 10621 cactgctgca cctggcccag cccgccaccc tctcgcagac catagtgccc atctgcctcc 10681 cggacagcgg ccttgcagag cgcgagctca atcaggccgg ccaggagacc ctcgtgacgg 10741 gctggggcta ccacagcagc cgagagaagg aggccaagag aaaccgcacc ttcgtcctca 10801 acttcatcaa gattcccgtg gtcccgcaca atgagtgcag cgaggtcatg agcaacatgg

## Figure 12 B (continued)

10861	tgtctgagaa	catgctgtgt	gcgggcatcc	tcggggaccg	gcaggatgcc	tgcgagggcg
		gcccatggtc				
		gggctgtggg				
11041	acctcgactg	gatccatggg	cacatcagag	acaaggaagc	ccccagaag	agctgggcac
11101	cttagcgacc	ctccctgcag	ggctgggctt	ttgcatggca	atggatggga	cattaaaqqq
11161	acatgtaaca	agcacaccgg	cctgctgttc	tgtccttcca	tccctcttt	gggctcttct
11221	ggagggaagt	aacatttact	gagcacctgt	tgtatgtcac	atgccttatg	aatagaatct
		agcaactctg				
11341	agctgtgtgt	gttgaggggg	atactctgtt	tatgaaaaag	aataaaaaac	acaaccacqa
11401	agccactaga	gccttttcca	gggctttggg	aagagcctgt	gcaagccggg	gatgctgaag
11461	gtgaggcttg	accagettte	cagctagccc	agctatgagg	tagacatgtt	tagctcatat
11521	cacagaggag	gaaactgagg	ggtctgaaag	gtttacatgg	tggagccagg	attcaaatct
11581	aggtctgact	ccaaaaccca	ggtgcttttt	tctgttctcc	actqtcctqq	aggacagctg
		gctcagtgtg				
		ggttcagccc			555 5 5	J <b>J</b>

## Figure 13 (A)

#### SEQ ID NO:3

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagccctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tetecactat ggeaactaae teacteaatt tteeeteett ageageatte catetteeeg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet aetetateae aaggeeagta eeacaeteat gaagaaagaa eacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta tteetcaate ttttacettt tecaaateec aateeccaaa teagtttte tetteetae teeetetee ccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccatte taagagettg tatggttatg gaggtetgae taggeatgat tteaegaagg caagattggc atatcattgt aactaaaaa gctgacattg acccagacat attgtactct ttetaaaaat aataataata atgetaacag aaagaagaga accgttegtt tgcaatetac agetagtaga gaetttgagg aagaattcaa cagtgtgtet teageagtgt teagagecaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagtigte etittetggt tiegigtiea ccatggaaca tittgattat agttaateet tctatcttga atctt

#### SEQ ID NO:76

ggcetctg actaactaat cactttccca tettttgtta gatttgaata tatacattet atgatcattg etttteet tacagggga gaattteata tettacetga geaaattgat tagaaaatgg accactaga ggaatataat gggattaggaa attacaggca tettecaacta ggeaactaac tacetcaatt tecetcett agcagacgate eacettettg ettetecaac caaaacatca atgttatta gtettata gtettata geagacgate eacettettg ettetecaac caaaacatca atgttatta gtetgaaagaa cacaggagga tettacett tecaaaceca accecaaa etetettte tecaaaceca aacetcat gaagaaagaa cacaggagga etttaceetttaceetttacee tecatggeg taaaagaacat eaceteette tecaaaceca acteettte eetttacee tecatggeg taaaagaaga gatggggage atcattetgt tatacetet tecaaacecag acttgettee ataagtetat tatatatat aaaatatata aatatatata aatatat

## Figure 13 (B)

#### SEQ ID NO:77

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagccctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tctccactat ggcaactaac tcactcaaat ttccctcctt agcagcattc catcttccg atettetttg etteteeaac eaaaacatea atgtttatta gttetgtata eagtacagga tctttggtct actctatcac aaggccagta ccacactcat gaagaaagaa cacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta tteeteaate ttttacettt tecaaateec aateeccaaa teagtttte tetteettac teetetete cettttacce tecatggteg ttaaaggaga gatggggage ateattetgt tataettetg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccatte taagagettg tatggttatg gaggtetgae taggeatgat tteacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagtigte cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

#### SEQ ID NO:78

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cecagecett gacaaaattg tgaagttaaa ttetecaete tgtecateag atactatggt tetecaetat ggcaactaae teaeteaatt tteeeteett ageageatte catetteeeg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet aetetateae aaggeeagta eeacaeteat gaagaaagaa eacaggagta gctgagaggc taaaactcat caaaaacact actccttttc ctctacccta ttcctcaatc ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctcc ccttttaccc tccatggtcg ttaaaggaga gatggggacg atcattctgt tatacttctg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgctttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgigtgtg tgtatgcgig igtgtagaca cacacgcata cacacatata aiggaagcaa taagccatte taagagettg tatggttatg gaggtetgae taggeatgat tteacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

## Figure 13 (C)

#### SEQ ID NO:79

#### SEQ ID NO:80

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagccctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tetecaetat ggeaactaae teaeteaatt tteeeteett ageageatte eatetteeeg atttettetttg ettetecaae caaaacatea atgtttatta gttetgtata cagtacagga tetttggtet aetetateae aaggeeagta eeacaeteat gaagaaagaa caeaggagta gctgagaggc taaaactcat caaaaacact actccttttc ctctacccta ttcctcaatc ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc cettttacce tecatggteg ttaaaggaga gatggggage atcattetgt tataettetg tacacagtta tacatgicta tcaaacccag actigcitcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

## Figure 13 (D)

#### SEQ ID NO:81

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagcctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tctccactat ggcaactaac tcactcaatt ttccctcctt agcagcattc catcttcccg atettetttg ettetecaac caaaacatca atgtttatta gttetgtata cagtacagga tetttggtet aetetateae aaggeeagta ceacacteat gaagaaagaa cacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta tteetcaate ttttaccttt tecaaateee aateeceaaa teagttttte tetttettae teeetetete ccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactet ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctgctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

#### SEQ ID NO:82

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg cttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagcctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tetecaetat ggeaactaac teaeteaatt tteeeteett ageageatte catetteeeg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet actetateae aaggeeagta ceacacteat gaagaaagaa cacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta tteeteaate ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc cettttacce tecatggteg ttaaaggaga gatggggage atcattetgt tatacttetg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttatcct tctatcttga atctt

## Figure 13 (E)

#### SEQ ID NO:83

gcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagcctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tctccactat ggcaactaac tcactcaatt ttccctcctt agcagcattc catcttccg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet actetateae aaggeeagta ceacacteat gaagaaagaa cacaggagta gctgagaggc taaaactcat caaaaacact actccttttc ctctacccta ttcctcaatc ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctccccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgctttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

# Figure 14

AATTOTGTA AGCATTTOOT ATGTGTACCT GCCCCTGGGC AAGGTGGGCC TGACTTGTTA	-1403
AGTGTTAGA GTTTTACCCT GTTCCTCTAG GAGGGCCTGG TACCACCACA GCCCAGCATG	-1343
TGTGGTGCC TCAGCAGGAG GCATCTGGTT ACAATCAACA CAAGCTGTTC CAGCCAATTT	-1283
NAGARACTI CAGGAGGRAT ACCGTTTTAG GRGGGGRTGG GGRCCCTCCT GCRCCCGRAG	-1223
CAGGATGTG CCACCAATCA TAAGGAGGCA GGGGCCTCCT TCCGCTGCTC CCTGGGACTC	-1163
PETAGGTGTC CGTGGCCTCA GCCCCCCTCT GCACACCTGC ATCTTCCTTC TCATCAGCTT	-1103
CTCTGCTTT AAGCGTAAAC ATGGATGCCC AGGACCTGGC CTCAATCTTC CGAGTCTGGT	-1043
acttanggig tactgacagt gigagaccct actcctctga tcaatcccct gggttggiga	-983
TTCCCTGTG CAATCAATGG AAGCCAGCGA GGCAGGGTCA CATGCCCCGT TTAGAGGTGC	-923
rgactiggag aaggaacetg ggcaagtcit cccaggaaca ggtagggcag ggaggaaagg	-863
GGGGCATCTC TGGTGCAGCC CGGTTCGGAG CAGGAAGACG CTTAATAAAT GCTGATAGAC	-803
TGCAGGACAC AGGCAAAGGT GCTGAGCTGG ACCCTTTATT TCTGCCCTTC TCCCTTCTGG	-743
CACCCCGGCC AGGAAATTGC TGCAGCCTTT CTGGAATCCC GTTCATTTTT CTTACTGGTC	-683
CACAAAAGGG GCCAAATGGA AGCAGCAAGA CCTGAGTTCA AATTAAATCT GCCAACTACC	-623
AGCTCAGTGA ATCTGGGCGA GTAACACAAA ACTTGAGTGT CCTTACCTGA AAAATAGAGG	-563
TTAGAGGGAT GCTATGTGCC ATTGTGTGTG TGTGTTGGGG GTGGGGGATTG GGGGTGATTT	-503
GTGAGCAATT GGAGGTGAGG GTGGAGCCCA GTGCCCAGCA CCTATGCACT GGGGACCCAA	-443
ANAGGASCAT CTTCTCATGA TTTTATGTAT CAGAAATTGG GATGGCATGT CATTGGGACA	-383
GCGTCTTTTT TCTTGTATGG TGGCACATAA ATACATGTGT CTTATAATTA ATGGTATTTT	-323
AGAITTGACG AAATATGGAA TATTACCTGT TGTGCTGATC TTGGGCCAAAC TATAATATCT	-263
CTGGGCAAAA ATGTCCCCAT CTGAAAAACA GGGACAACGT TCCTCCCTCA GCCAGCCACT	-203
ATGGGGCTAR ARTGRGACCA CATCTGTCAR GGGTTTTGCC CTCACCTCCC TCCCTGCTGG	-143
ATGGCATECT TGGTAGGCAG AGGTGGGCTT CGGGCAGAAC AAGCCGTGCT GAGCTAGGAC	-83
CAGGAGTGCT AGTGCCACTG TTTGTCTATG GAGAGGGAGG CCTCAGTGCT GAGGGCCAAG	-23
CANATATITE TEETTATEGA TTAACTCEAA CTCCAGGCTE TCATEGCGGC AGGACGGCGA	+38
ACTIGCAGTO TOTOCACGAC COGCCCCTGT GAGTCCCCCT CCAGGCAGGT CTATGAGGGG	+98
TETEGAGGGA GGGCTGCCCC CGGGAGAAGA	
1350 bp	
MPP TED GIN LE	u -39

MET TRY GLM LEU -39
----- AAGAAGTCCT CCTCAGACAG GTGCCAGTGC CTCCAGAATG TGG CAG CTC +1527

Figure 15

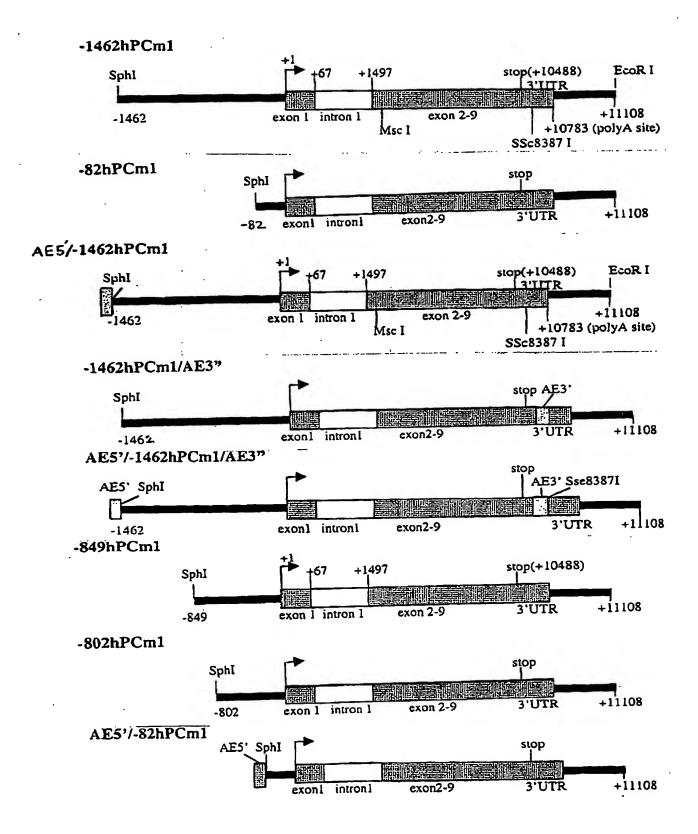


Figure 16

hPC Minigene	Constructs	Expression Activity (%±SD)
-1464hPCm1	PA	100
-82hPCm1 -82	PA	98.7±11.8
AE5'/-1464hPCm1 AE5 -1464	* PA	101.9±12.5
-1464hPCm1/AE3"	* AE3 pA	70.1±7.5
AE5'/-1464hPCm1/AE3"  AE5  -1464	* AE3 pA	74.0±3.8

Figure 17A

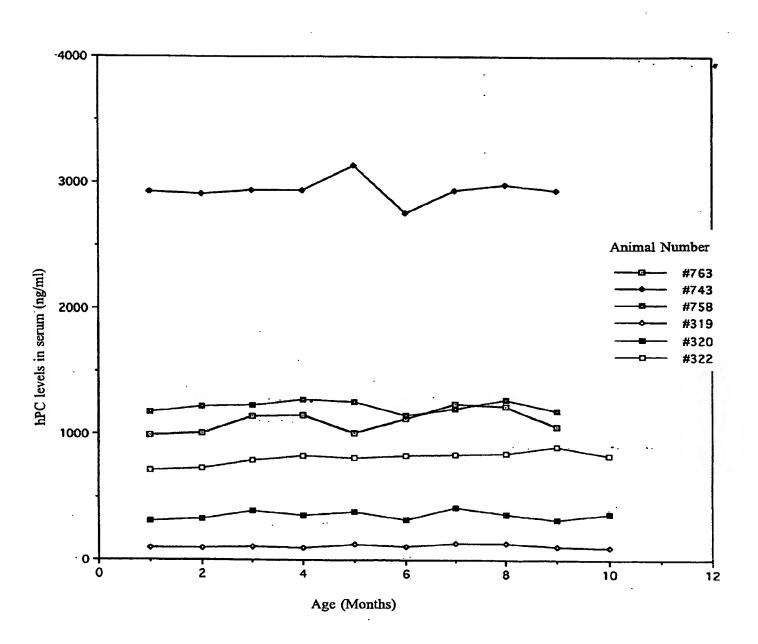


Figure 17B

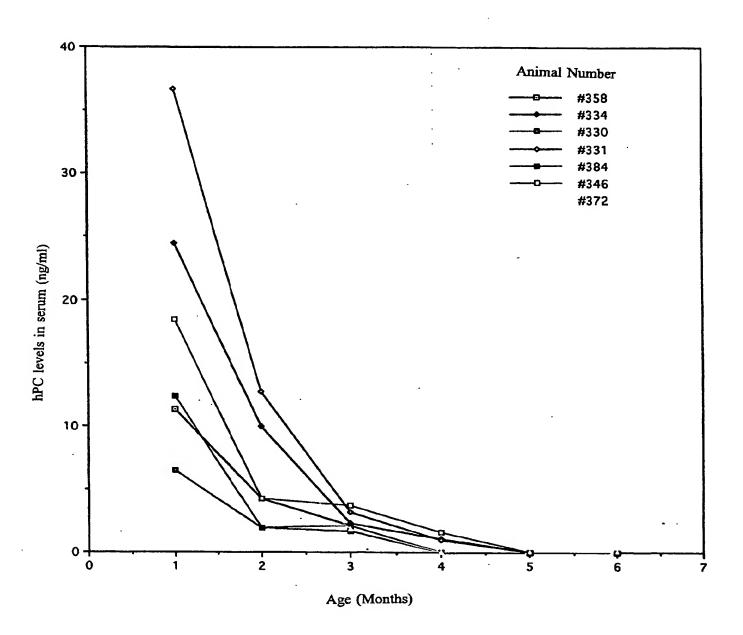
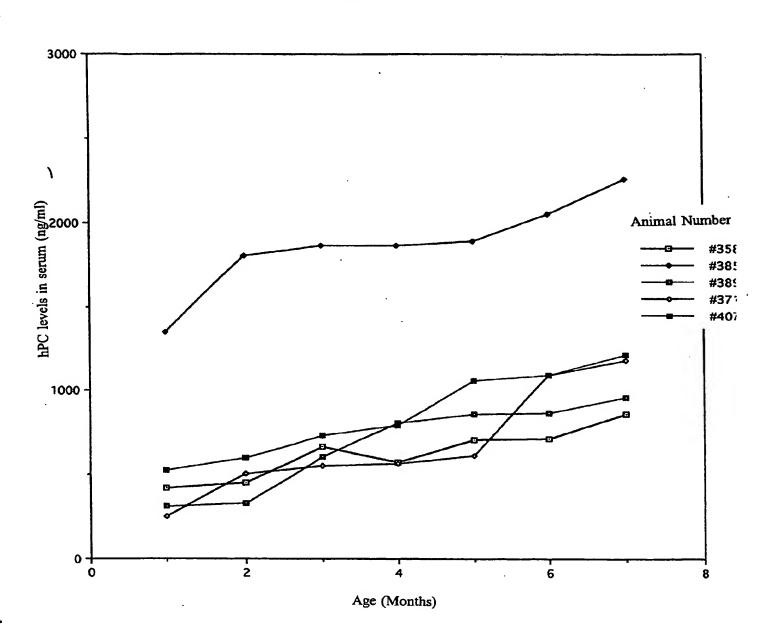


Figure 17C



### SEQUENCE LISTING

5	Kurachi, Kotoku  Kurachi, Sumiko	
	<120> Nucleotide Sequences for Gene Regulation and Methods of Use Thereof	
10	<130> UM-04305	
10	<140> xx	
	<141> 2000-06-06	
	<150> 09/328,925	
15	<151> 1999-06-09	
	<160> 208	
20	<170> PatentIn Ver. 2.0	
	<210> 1	
	<211> 32	
	<212> DNA	
	<213> Homo sapiens	
25		
	<400> 1	
	agccattcag tcgaggaagg atagggtggt at	32
	<210> 2	
30	<211> 7	
	<212> DNA	
	<213> Homo sapiens	
2.5	<400> 2	
35	gaggaag	7
	<210> 3	
	<211> 1273	
	<212> DNA	
40	<213> Homo sapiens	



<400> 3 ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60 gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120 gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180 5 cagecettga caaaattgtg aagttaaatt etceactetg tecateagat actatggtte 240 tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tcttcccgat 300 cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360 tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420 tgagaggcta aaactcatca aaaacactac tccttttcct ctaccctatt cctcaatctt .480 10 ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctcccc 540 ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600 cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660 aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720 aagttatttt atatatata tatatatata aaatatataa tatacaatat aaatatatag 780 15 tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840 agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900 agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960 ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020 ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080 20 aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140 ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200 aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttaatccttc 1260 tatcttgaat ctt 1273 25 <210> 4 <211> 38059 <212> DNA <213> Homo sapiens 30 <400> 4 gtatatctag aaaaccccat tgtctcattc caaaatcacc ttaagatgga taggcaactt 60 cagcaaagtc tcaggataac aaaatcaatg tgcaaaaatc acaggcattc ttatacacca 120 atagcagaca aacagacagc caaatcatga gtgaactccc attcacaatt gcttcaaaga 180 gaataaaata cctaggaatc ctacttacaa gggatgtgaa ggacctcttc aaggagaact 240 35 acaaaccact gctcaatgaa ataaaagagg atacaaacaa atggaagaac attacatgct 300 catgggtagg aagaatcaat atcatgaaaa tggccataat gcccaaggta atttatagat 360 tcaatgccat ccccatcaag ctaccaatga ctttcttcac agaattggaa aaaactactt 420 taaagttcat atggaaccaa aaaagagccc gcatcgccaa gtcaatccta agccaaaaga 480 acaaagctgg aggcatcatg ctacctgact tcaaactata ctacaaggct acagtaacca 540 40 aaacagcatg gtactggtac caaaacagag atacagacca atggaacaga acagagccct 600 cagaaataat gccacatatc tacaactatc tgatctttga aaaacctgac aaaaacaaga 660 aatggggaaa ggaatcccta attaataaat ggtgctggga aaactggcta gccatatgta 720 gaaagctgaa actggatccc ttccttatac cttatacaaa aattaattca agatggatta 780 aagacttcat tgttagacct aaaaccataa aaaccctaga agaaaaccta ggcaatacca 840

ttcaggacat aggcatgggc ttggacttca tgtctaaaac accaaaagca atggcaacaa 900

45



	aagccaaaat	tgacaaatgg	gatcaaatga	aactaaagag	cttctgcaca	gcaaaagaaa	960
	ctaccatcag	agtgaacagg	caacctaaag	aatgggagaa	aatttttgca	atctactcat	1020
	ctgtccaagg	gctaatatct	agaatctaaa	atgaactcaa	acaaatttac	agaaaaaaac	1080
•	aaacaacccc	atcaacaagt	gggtgaagga	tatgaacaga	cacttctcaa	aagaagacat	1140
5	ttatgcagcc	aacagacaca	tgaaaaaatg	ctcagcatca	ccggccatca	gagaaatgca	1200
	aatcaaaacc	acaatgagat	accatctcac	acaagttaga	atggcgatca	tcaaaaactc	1260
	aggaagcaac	aggtgctgga	gaggatgtgg	agaaatagga	acactttgac	actgttggtg	1320
	ggactgtaaa	ctagttcaac	cattgtggaa	gtcagtgtgg	cgattcctca	gggatctaga	1380
	cctagaaata	ccatctgacc	cagccatccc	attattgggt	atataccaaa	gtattataaa	1440
10	tcatgctgct	ataaagacac	atgcacacgt	atgtttattg	cggcactttt	cacaatagca	1500
	atgacttgga	accaacccaa	atgtccaaca	atgatagact	ggattaagaa	aatgtggcac	1560
	atatacacct	aggaatacta	ggcagccata	aaaagaaaat	gagttcatgt	cctttgtagg	1620
	gcatggatga	agctagaaac	catcattctc	agcaaactat	cgcaaggaca	aaaaaccaaa	1680
	caccgcatgt	tctcactcat	aggtgggaac	tgaacaatga	gaacacttgg	acacaggaag	1740
15	gggaacatca	cacaccgggg	cctgttgtgg	ggtggggggc	gaggggaggg	atagcattag	1800
	gggatatacc	taatgctaaa	tgacgagtta	atgggtacag	cacaccaaca	tggcacatgt	1860
	atacatatgt	aacaaacctg	ctcgttgtgc	acatgtaccc	taaaacttaa	agtataataa	1920
	taaaaaaaag	atcattctaa	aatttataca	agcccttaga	acagttaaaa	atatcttacc	1980
	aaaagaagaa	taaagttgga	ggaatcactc	tacctaatat	aaagtcttac	tacatagcta	2040
20	cagtaattat	gacagtgtta	tattggcaga	gggataaata	catcaatggc	acaaagaata	2100
	gatagagaaa	ctggaagtag	acccaaaaca	atatggttaa	ctgacttacg	aaaaaatttc	2160
	agaagccatt	cagtcgagga	aggatagggt	ggtattgttg	ttttttgttt	taacaaattg	2220
	tgctggataa	attggacata	cctatggaaa	aaaaaatgaa <sup>.</sup>	gtttgaccta	aacatcatac	2280
	tttacacaaa	tattaactca	aaatggagca	tgggcataaa	tctaaaactt	caaactgtaa	2340
25	aacatttaga	aaaaaatagg	aaaaaacta	tcaggatcta	gtgttagtgg	aagagttcta	2400
						aaaattctac	
						caatatattt	
	aatccacgta	tctatgaaag	gattcatatc	tagaatatat	aaacaacctt	aagaatctga	2580
20						gaaatgtaaa	
30						gggtcccact	
						tgaagatgtg	
		•		•		atcagaggtg	
						ttggaaatag	
25		_		_	_	gctcagcttg	
<b>35</b> .				•		ggttatgcag	
					•	tttaggatat	
						gtatgcttgc	
					_	acatgatttg	
40						gttctttgtt	
+∪						ttatgatgca	
						gttagaaaac	
						cagaatatag	
						aaaatgggaa	
45						tatatgacat	
70	LLLaayyLLE	cayaaytatg	caacyayycg	cycecedat	LLLLLAddEE	atatatcttc	2000



	aatttaaagt	tttagttaaa	acataaagat	taacctttca	ttagcaagct	gttagttatc	3660
	accaaagctt	tţcatggatt	aggaaaaaat	cattttgtct	ctatgtcaaa	catcttggag	3720
	tttgattatt	tggggaaaca	caatactcag	ttgagttccc	taggggagaa	aagcaagctt	3780
	aagaattgac	ataaagagta	ggaagttagc	taatgcaaca	tatatcactt	tgtttttca	3840
5	caactacagt	gactttatgt	atttcccaga	ggaaggcata	cagggaagaa	attatcccat	3900
	ttggacaaac	agcatgttct	cacaggaagc	atttatcaca	cttacttgtc	aactttctag	3960
	aatcaaatct	agtagctgac	agtaccagga	tcaggggtgc	caaccctaag	caccccaga	4020
•	aagctgactg	gccctgtggt	tcccactcca	gacatgatgt	cagctgtgaa	atcagactga	4080
	aatgctgaaa	taacgataaa	aaaaaataca	gaggttaaac	tagcaaagtg	agtaaagtca	4140
10	agggataaag	aaaatttgtt	ggaaaactca	caaagcagga	cataaagcaa	ggccattaga	4200
	tatatctcat	tagtgtgaca	tctgggagga	caaagcatcc	aaaccctttc	ttctatataa	4260
	gtggtgagat	gatgaaggtt	gtaagaggct	tetgeceet	tgaagacttc	agatgctggg	4320
	gaaaggatag	ataagaataa	ggatgaacct	ggcttttgga	gcctgggaaa	taatgactag	4380
	cgataaacct	gaagggaagt	taagtatacg	atccccagat	aatactaagg	agaaaggcaa	4440
15	tgtgattctg	cagccattgt	agccagagat	aataagccct	tgaggaaggg	gccaggggaa	4500
	tttttctaag	gatagacagt	attaatgcag	cactctcttc	tgctattaaa	ctctcattgg	4560
	cttctaaaag	gagtttcggt	gagtgatttg	ctgagatgtt	tgcattttca	tgctgctgcc	4620
	tttaggttat	tattgcaaca	gtttggaatt	ttgaaattaa	aacagttctg	taaaaccagt	4680
	ttagttttgt	aaagtgtatg	catcaaagat	gtccttcatt.	cagacattac	tgagttacaa	4740
20	ctacggtgcc	aggtactgtg	tcagggtact	aggggtatgg	ggataaacca	gactccctct	4800
	ttgatctaaa	gcagcatgag	gccaggtgag	aggtttcaat	atatgtgata	aaatgtgcac	4860
	taggtactaa	gggatcatag	agaaaggaac	acattaaatg	gggaaacaat	tgatagagag	4920
	agaatatttt	catctgggtc	ttaaaagatg	agtaggcgtt	ctctctctt	aaatgtctga	4980
	tataagggca	ttttatgcaa	agaaggatca	ctcgtgcaaa	gactcagctt	tgcaagaacg	5040
25	tgaggtattt	caggagtttt	gtatggttcc	atatggacta	tgacaagtga	gacaggtaaa	5100
	ctaggcagag	ctggtcatca	gataatgaag	tcattaacct	aaggagattg	gacaataaaa	5160
	tgcaatatgg	aggtatcgaa	gtataaacat	aaggagtacc	actgatggct	gatttaggat	5220
		gcaacacgct					
••		tcctgagact					
30	•	tacaaaaaaa					
		tčtgggtatg					
		gaccacagaa					
		tgaatgtttt					
2.5		ttataaaaat					
35		atggatgcaa					
		ctattactgg					
		gagacgtatc			_	_	
		gcacgtgtgt					
40		gcagtagtgg					•
40		ggtaagcaag					
		aaatattgtg					
		ataaggtgaa				_	
		ccatggagag				-	
15		aaagcgggga 					
45	agggagagat	caacaattag	aaattattta	gagcagccag	taaatacata	aactcaaaat	6300



WO 00/75279



	tattctttag	gtcattcctg	attgtgacaa	tagtcatttc	attatataaa	tgtgattaag	6360
	gaaggaaaag	agctacacag	aagttattaa	agagctaaag	agaattgaga	aatttaaaac	6420
	agaagaaagt	agggccaaca	tgaaaggagt	agggagaaaa	agagataacc	agcatattgt	6480
	ctcactgatc	ctgccaacac	ctgtgagata	gatattctta	ttatactact	gttaaaccta	6540
5	atttacattt	gacaaagtta	aggttcagag	cttgtgtgac	ttgtccaagg	tcacaggtct	6600
	agaggaggca	gatacttgat	tcaaacctat	ttctgtctga	tctgattcta	aagtctgttt	6660
	tttcactcaa	ccacactgta	cagtcagctc	tccttgtgag	ttccacagcc	acagattcaa	6720
ı	ttaactgcag	atcaaaaata	ttcaagaaaa	aaatggatgg	ttcgatctct	actgaacatg	6780
	tacagactct	tttatctttc	attattccct	aaacaataca	gcataacaac	tatttacata	6840
10	gcatttacat	tgtattagct	attaagagaa	acctagagat	gatttaaagt	acaaaggagg	6900
	atgtgtttag	gttatatgca	aatagtaagc	catttttata	tcggagactt	gagcatccac	6960
	agatcttgat	atttgcaggg	ggtcttgcca	ccaattttcc	atggatactg	aggaacgact	7020
	gtaaatggat	gcaggcatgg	atgctattta	ggagtgtcca	gggccaagta	aatgagttgc	7080
•	tgagcagaga	ggtgggtgga	ggctgtgagg	catcaatatg	tggtggcatc	catctgcatt	7140
15	ttggtgattt	ttttccttca	cagtcctcgg	ctgtctggga	agagaaggat	gaaggcagat	7200
	ggctgctcca	atttaggggc	taggattgca	gggtgggcac	agcattgcaa	acgagtgaag	7260
	gaaattgaga	aatatggcca	atgaagagtt	gaagagaggc	ctggcatggt	ggctcacacc	7320
	tataatccca	gcactttcag	aggcccaggc	aggcagatca	cttgaggtca	ggagttcgac	7380
	accagcctgg	ccaacaaggt	gaaatggtga	aaccccggct	ttactaaaaa	tacaaaaatt	7440
20	agctgggcat	ggtggcgggt	gcctgtaatc	ccagctactt	gggaggctga	ggcaggagaa	7500
	tagcttgaac	ctgggagatg	gaggttgcag	tgagctgaga	tcgcaccact	gcactccagc	7560
	ctgggcgaca	gagcaagact	ctgtcaaaaa	aaaaagagtt	gaagagaaaa	agtctaggct	7620
	aaattcaaag	aaaaaagtg	agcccaaaag	gaacttgcag	agcaagggaa	aagcagggat	7680
	gtcaagggac	tagaacactc	cataaagtga	acagctgcaa	tgaaaataag	ggaagaaagt	7740
25	ttagttcatc	tccgtttctt	tcctttcctt	tttactttcc	tttctcttcc	tttttggagt	7800
	tagtcaggaa	gtagtcccaa	ataccccaga	aagttcatct	tataagccct	tggtcctctt	7860
	gagatggtat	cagatatatt	gctagaccct	tgaagaaagg	aacaactcca	ggcaacttct	7920
	tgagtccctg	ttattaattt	tatacataca	cacacatata	tgtatataca	tgaaaacaca	7980
	caaacacatg	tgtgtgtata	cagccatgca	ttccttaaca	atggggatat	attctgagaa	8040
30	atgtgtcatt	aagcaatttc	atcattgtgc	gaacataata	gagtgtactt	acctaaacct	8100
	aaatggtata	gcttactaca	tacctaggtt	gtattgatgt	ggcctattgc	tcctaggctc	8160
	ctgggctgca	aacctgtaca	gcatgtgact	gtactgaaca	ctgtaggcaa	tggtaacagt	8220
				- <del>-</del>	_	taaccttatg	
2.5						agcaaatgat	
35						tgtatatatg	
						tatgtacaca	
	•				-	tatatgtata	
						gagagaga	
40						gagacatcta	
40					_	aaaaacatat	
•				•		ccaaccaaag	
٠			•			taacatctag	
				-		tcttcaggag	
15						agggtgatgg	
45	atcactttgc	aaagatcctc	aatgagctat	tttcaagtga	tgacaaagtg	tgaagttaac	9000

	cgctcaattt	gagttctttc	tttttcatcc	aaagtaaatt	caaatatgat	tagaaatctg	9060
	accttttatt	actggaattc	tcttgactaa	aagtaaaatt	gaattttaat	tcctaaatct	9120
	ccatgtgtat	acagtactgt	gggaacatca	cagattttgg	ctccatgccc	taaagagaaa	9180
	ttggctttca	gattatttgg	attaaaaaca	aagactttct	taagagatgt	aaaattttca	9240
5	tgatgttttc	ttttttgcta	aaactaaaga	attattcttt	tacatttcag	tttttcttga	9300
	tcatgaaaac	gccaacaaaa	ttctgaatcg	gccaaagagg	tataattcag	gtaaattgga	9360
	agagtttgtt	caagggaacc	ttgagagaga	atgtatggaa	gaaaagtgta	gttttgaaga	9420
	agcacgagaa	gtttttgaaa	acactgaaag	aacagtgagt	atttccacat	aatacccttc	9480
	agatgcagag	catagaatag	aaaatcttta	aaaagacact	tctctttaaa	attttaaagc	9540
10	atccatatat	atttatgtat	gttaaatgtt	ataaaagata	ggaaatcaat	accaaaacac	9600
	tttagatatt	accgttaatt	tgtcttcttt	tattctttat	agactgaatt	ttggaagcag	9660
	tatgttggta	agcaattcat	tttatcctct	agctaatata	tgaaacatat	gagaattatg	9720
	tgggttttt	ctctgcataa	atagataata	tattaaactt	tgtcaaaagg	actcagaaag	9780
	atcagtccaa	ccctctaacc	catattggat	ggtgatatac	tacagggtta	tgccagtgtg	9840
15	ggaactatcg	ctggtaaata	agtttaatcc	tccctagggc	ttcacaaaga	acattgttcc	9900
	accccaggag	ggtggaagga	agaaactgaa	atgattgtgt	cttagaacct	aatgaaagtt	9960
	tgcattcctc	agtaaaatca	gagactgctg	attgacttaa	atgtttatag	cttcaaagtc	10020
						ctccccatta	
••						accctggcct	
20						tctggctgtt	
						ttcgggtctt	
						cttcaacctg	
•						tttaacagag	
25		<del>-</del>				tttattttct	
25						ttgtttattg	
						tatttatcaa	
						taaagctaca	
						tgacttacct	
30						aacttccatc	
30						tggacactct	
						gaaaaagagt	
						actaccatac	
					-	tctccattcc	
35						atgccgattt tcactccca	
33						gtacactgtt	
	•					ttactactgc	
					_	cctqcctctt	
			_	_		tcttgcacat	
. 40						tcacataaag	
•						ctttccagtt	
ĺ			•			attattaact	
•					_	aaaggtctcc	
					-	aaccccttgc	
45						aacaggaacc	
•		,	J				

		tggtagtctt	aaacaggaat	ctctcacttg	gtaagatctt	gtctcttgtt	gtatttgacc	11760
		ccaactgtct	atggctttgc	ctgaacccaa	agtacacaca	gcctagaaac	caaaggagaa	11820
		ccaaatgtgg	gataaaatga	cactcatttt	aacgacatgt	ctcagcaaat	gagttcctgt	11880
		gtagctggct	gaaagcccag	accctttcag	taaaacatcc	tgaataattç	acatttgttg	11940
	5	gtctataata	taaagggcaa	atgtagctca	tttttagacc	agttctgaac	atcaatagta	12000
,		acaaaccaga	gataaccgat	tttgttttca	tagaattgga	acaaattaga	gtatctgtgc	12060
		aaaagcatat	cagatctagg	agcagagggg	acaaggtcta	atttttaaat	aagcaaattt	12120
		tccagggagg	gactacttat	gataaaggga	tattagtctc	ttagtcaacg	gaacctggat	12180
•		acacgcttct	gacagagaag	agggagaata	ggcaggaatc	tacacaccag	atgtcaagga	12240
	10	gatttgcttt	aaaatacgac	tgataattag	aaatttctca	gtttccccct	tttccctcat	12300
		tctttgattc	ttattgttat	ctttatctct	tactcctttg	tttctcatat	attgagtctt	12360
		acagatcaag	ctcccatttt	tttcttcagg	ggtattttc	tagttcaaag	tgcctaccat	12420
		ctcccttctg	gttctattca	tccttctctc	ccaaagctcc	tttagaagtg	tggattaagg	12480
		cagagcacta	agaaaccaga	cttaaagatt	cccttctcat	tctgactttt	ctcctttcac	12540
	15	ctattccttc	ctcctgtttt	cttaccatca	gtgtcttcaa	aggctttcaa	gtacacggta	12600
		aatgcagaaa	cttcaagaaa	ggcagaatgg	aaacataacc	aatgcataca	taaataaagc	12660
		acactgtaga	atctttttaa	attctgtatg	atatatcgaa	tgctgtctct	cacattacct	12720
		agaccatttg	aaaccgaatt	tgtaaaacat	agactatctt	taagtagtaa	cagatgette	12780
	• •	tgacatgttt	tctattgtct	tgaaccatta	ctgcatatga	tacatcaaag	ttaagtgaca	12840
•	20					cagttcctaa		
						atagagcctt		
						taactctgcc		
						gtacctttct		
	25					atgccttcct		
	25					gcttttagaa		
						cagtagtttt		
						gaccgggcat		
						agatcagtgt		
	20					ctatgaatgt		
	30					attttttgaa		
						aatattggtg		
						ttgtagtgat		
						ttggtaacta		
	35					ccgtatgtgg	_	
	33					aacggccaaa		
						aaaacccatg		
						aaatcatgac	=	
						tgactggcca		
	40					taaacttgtt		
•	40					aatgttacat		
						tctttttccg		
						ttattacctc		
•						gaggaattca		
	45					tttaacacca		
	73	yactaayyat	aaayagtaga	aaactagcag	gactctatta	accgattaca	gcaatccacc	14400

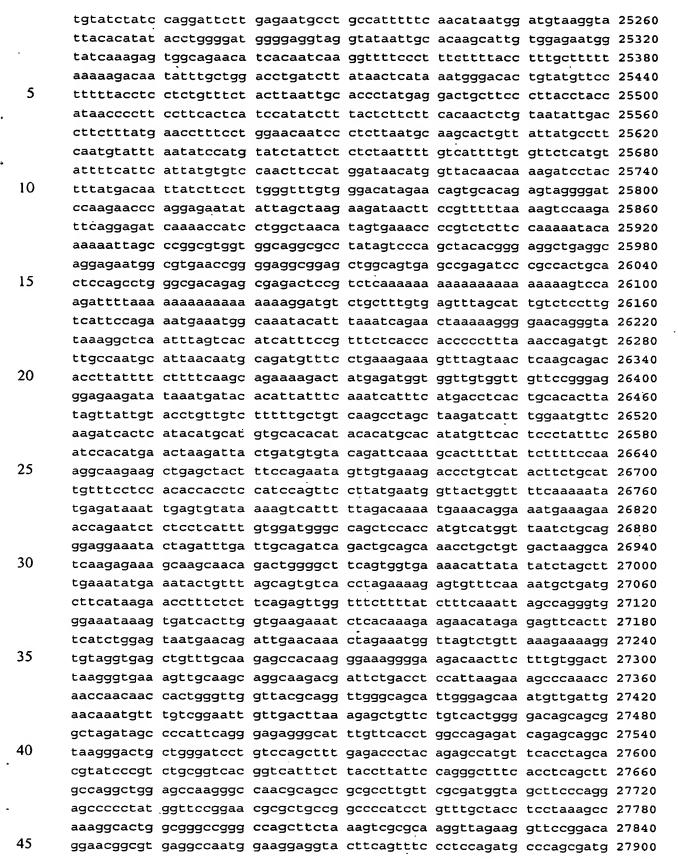
	tgacagatga	aaaaggcatg	aaatgaaatg	aaatgtagca	gctacactcg	tcctattgag	14460
	aaaggaaaaa	agtcacctgt	aatgttgttc	agaaatcctt	tcagtacțaa	aaaattcatt	14520
	gaccatcttc	ctttagtctc	gaaaatttct	tagaaggtaa	aaaaaggaaa	aggtgacagg	14580
	gcaaagacat	ttgaaaagaa	agaaaagagt	gaatgaactt	gcacacctgg	cttggactcc	14640
<b>5</b> '	ccattcccct	taggtttcca	ttgtggggga	caaactaatg	cctgggttac	ctttcttgag	14700
	agtgtgttaa	ttgattcaat	atctctgaag	tgctactttc	atctgaaagg	ttataatttg	14760
	aaattcagat	ttacctggat	aaatttgatc	ttgctattat	ggaaacctct	agaaatcctt	14820
	ggagtagtta	ctcattatca	gcttaaataa	tatagccggt	ggagctgagg	gaatgagtaa	14880
	ctcaattagt	ctcagttaca	actgaagggc	acattgttgt	aaactataat	tgaaaacata	14940
10	aatatcttta	cctagtttaa	aaaataaaga	tgctttaaaa	ggaggaaggg	aatagccctg	15000
	aggaatgtaa	atataagcac	aaaacttcta	caacagagtt	tgctacgtgt	gtggctgtgt	15060
	tccacccagc	aaaaatgcta	agtctacaac	tgacacaact	tggatactct	catgttccca	15120
	cattttggtt	tggtcaaggc	tgtgcagttg	tactgcaggc	caccaccact	cctggcctct	15180
	acagtatatt	gatctgaccc	accaatctga	tcaaggttta	gaaaaatatt	ttcagcccag	15240
15	ttagctcaca	aacaaaatga	gaattcccac	aaattgctct	ttatctcaga	caacagaggá	15300
	aagctacagc	aaaagcataa	acaaattacc	atttaagttt	gttgcttcaa	attaaagact	15360
	aattgcaaca	getactagat	agcacagttt	atggggcatc	tcggccccaa	gtcttttgtc	15420
	ttataaggtc	ttgaaaaaaa	gaaaggagat	tttcatcaat	aagagttttt	tgttatcttt	1.5480
	ttcccttgtt	catcaggccc	ttcactgcga	gagagaggtg	taaacgttca	gggcatgcat	15540
20	tctagttaaa	gaatattaat	tggctattgg	gtccctttgg	ttagaataaa	gacctctgta	15600
	tgatgtccct	agctgtacat	caaacccaaa	tatctctcag	ataaatgaag	gtctgtaaga	15660
	•					gtaaagtaga	
						ccagcatgtc	
0.5						aaatatgttg	
25						tcttttgtcc	
						taaactaaag	
						gtgttaacat	
					_	gcactgagac	
						aatgctcagt	
30						actctcccct	
						agggcaaaag	
						agatcaacag	
						tacagaggtc	
35						tttattcatt	
33						acaatctcat	
						gcattgcttc	_
	•					actgttagtg	
						ccatcttttt	
40						gttcctttgt	
					_	gactaagagc ttagattcat	
						tgttcccca	
						aaaggaactg	
						ctactcgttt	
45				·		aattggcctc	
-				2222220000	Januageca	anceggeete	T, TOO

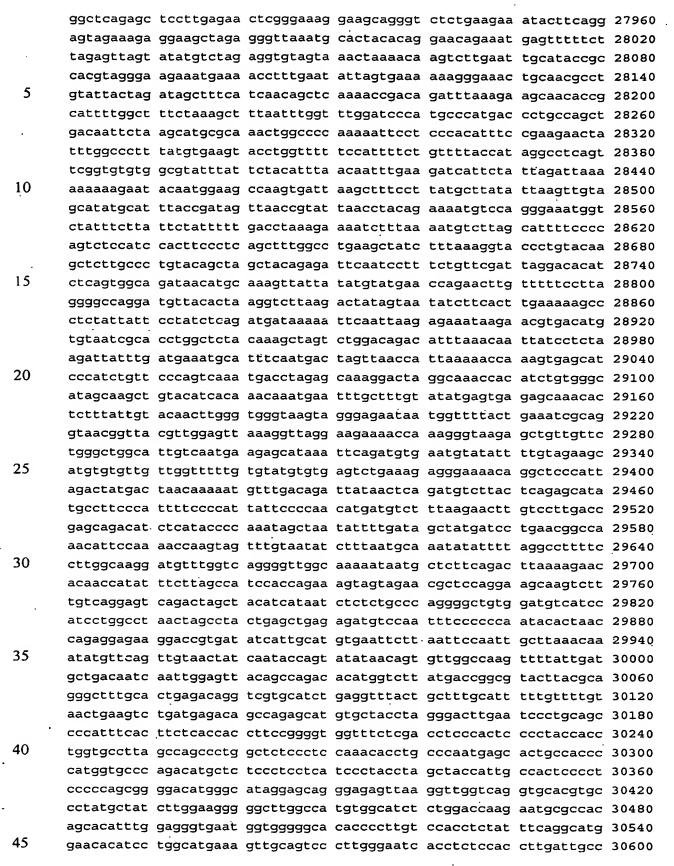
	aggttagtcc	acctgagtgg	gaagaagcgg	tgaaagaatt	tgtctgtcag	tatctgtctc	17160
	tcattggtta	gaagttcgac	ttatggggaa	ttaactccct	cacatttcct	agttggatag	17220
	cttgggtacc	agaggcatat	ggcatccatg	ctcagcatga	acagggaagc	ttcaaggcaa	17280
	aagacacata	gtgcagctat	gagccaaggc	aattcaagga	tacacccata	ggaggctggt	17340
5	tgacatccac	ccagagctaa	tcaccaccat	gctggaaaaa	gacacaggtg	aagctgagaa	17400
	gaatgaaggt	ggtgcatagg	aggtatctaa	tacagtcact	cattttcaaa	ctttccatgt	17460
	tatgattgca	ctgaccactg	aggatttcta	ttgaaagttt	tactgttgtc	aaacacgtac	17520
	acaaggggaa	aggtgtctta	cattgtttat	gttcctgtgc	tgctctagaa	acagaaatag	17580
	gctcaagagc	agagcctgtt	tttcttaatt	cagcaggtct	aagctaacaa	gtcctgaaac	17640
10	atggtacttc	ctgttattgg	tattgcatag	gagaaacaaa	gggaaagcac	agtaattaga	17700
	aaatacaaac	aagatggcag	gaataagcca	aaaatatcag	gaaacacaat	tattgtgaat	17760
	tgggattaaa	ctaatctatt	aataatgaca	actttcagct	tggagttaaa	aatttaattg	17820
	tatactgtta	acgaaagtga	tacctaaaat	aaaattacac	tgggaggcca	aaatgaaggg	17880
	atgtgaaaag	aactatcagg	taaaaactaa	caaaaagaaa	ctagcaaagc	aatcttaata	17940
15	tcagacaaaa	tagaatccaa	gaggaaaatc	atttcaaaag	acaagagatt	ttttttatta	18000
	ataaggggaa	ttgcatagga	gagtaaagaa	aatgtgggcc	actggaatgc	ttagcactaa	18060
	tgacatattg	gtctttggtc	ttcagttacc	ttacaggacc	ctatttcatt	ctcttatgtt	18120
•	tgatatgtaa	ccacctcagc	cagcttcaag	ttgctttttg	gccctaatgg	acttcctagc	18180
	actataattt	cttttttt	aaatgtttta	ttttaggttt	aggggtacat	gtgaaggttt	18240
20	gttacataga	taaacatgtg	tcacaggggt	ttgttgtaca	tattattaca	tgacgcagat	18300
					•	accetectee	
	ctcaagtaga	ctccagtatc	tgttgtttcc	ttctttgtgt	ttataagttc	ttaacactta	18420
					_	tagtttccta	
25						cttctttct	
25						gtctgtcatt	
						taaagacttc	
						aagcaatttc	
						tacatgagat	
20 .		_				gtaactttca	
30						taaataagta	
						tatataccag	
•						taataaatgt	
						aaaaattaag	
35						aatatagtaa	
<b>J</b> J						ggctttttg	
						aaatgcaaag	
						acactaccag	•
						cacaaggata	
40			•			atatataaag	
,,						gttttcaata	
•						aacaatttat	
						ccataaagga	
			•			aagacactac atacctaata	
45						agaaaattgc	
	account	acacagugat	Jacadage	caagettyta	caaccaccit	ayaaaattyc	13000

	ttggcagtat	ctactagatc	tgaacatgtg	atccagtaat	tacactcata	attataagcc	19860
		catgtttatg					
		ccaaaaactg					
		atagtaatac					
5 .	acaaaacaac	atggtttaat	ctcacagaca	aaatgttaaa	tgaaagacac	agacgagtac	20100
		cttctgttta					
	tccaggtaat	ggtaacctat	aaaaaggaaa	aagggtggaa	tgattgggag	ggggcatctt	20220
	ctggggtctt	gataatgtgc	tatgtattgg	tcagtttagt	gtttaaacag	gctcatttac	20280
	tttgtgaaaa	cttacactaa	aattgtgtgt	attttttgaa	tatatgttat	acattaataa	20340
10	atagggtttt	taaacctgta	gttcataatt	tagtgaaagt	agaatatcca	aacatttagt	20400
	tttaaaccaa	tcaattatag	tgctaccatc	atttttatgc	attattgaga	agtttattt	20460
	acctttcttt	ccactcttat	ttcaaggctc	caaaatttct	ctccccaacg	tatattgggg	20520
	gcaacatgaa	tgcccccaat	gtatatttga	cccatacatg	agtcagtagt	tccatgtact	20580
	ttttagaaat	gcatgttaaa	tgatgctgtt	actgtctatt	ttgcttcttt	tagatgtaac	20640
15	atgtaacatt	aagaatggca	gatgcgagca	gttttgtaaa	aatagtgctg	ataacaaggt	20700
•	ggtttgctcc	tgtactgagg	gatatcgact	tgcagaaaac	cagaagtcct	gtgaaccagc	20760
	aggtcataat	ctgaataaga	ttttttaaag	aaaatctgta	tctgaaactt	cagcatttta	20820
	acaaacctac	ataattttaa	ttcctacttg	aatctgcttc	cttttgaaat	catagaaaat	20880
	atcagtagct	tgaattagac	caattaattt	tctagattgc	atcatattt	aaatataact	20940
20 ·	atgtaatcat	ctacaacctg	aattctttct	gtgtccaatt	tgtccaattt	ttttctctaa	21000
	catttatatc	acaaagcaat	taatttgtgt	gatttctgca	tatgtatttg	taattcatca	21060
	agtcaaatca	atgtagtaat	actatatcat	aaaatataca	caaataattg	agtgataggc	21120
	ttctagtata	aggacggtaa	gtttgaagca	tgattctatc	tgggctggct	agtttactct	21180
	gagaaagtta	ttttttattg	ttgggtctta	agctgagttt	acacacttgg	tgtcagaatg	21240
25	attccggcaa	tgaactgttt	tatgttctgc	taggctgatc	agcacaatct	atatggctgt	21300
	gaacaaaaca	atgtttccca	gtcataccaa	ccatgccacc	attttaacag	ctgattagtg	21360
	tattcagaac	atctccactc	catgttcgta	tggctgttat	ctaaagatga	aagcagtaga	21420
•	cacttttatt	ttttgaaaaa	tttaggctct	gcagggtcaa	ttatatttga	taaatgaggg	21480
	gcttttttga	agcaaactag	atataatttc	ttttgcattt	ctaaagcctg	atatcttatt	21540
30	aattggtaca	ttaaattgtg	caccatttct	ctgtaactgt	ttcagtacct	gtctcagcac	21600
	tataccaggc	agaagaaata	aagaaagaac	cagtgccaga	tcagcttggt	caggagaccc	21660
		gcactagagg					
		gtctcacagc					
		gcaagccagt					
35		gctagttatc		-			
		ggtttaagaa					
		attctggtaa					
		tgctgcagag					
40		ctgttcccaa					
40		aaggetteag					
		gctagtctac					
		cttaaaaacc					
		tctatgctat					
15		gaatatagcc					
45	ttttctgcct	tttcaatgat	aaaggtccct	tttctgtagc	cattgttgat	tgtgtacact	22500

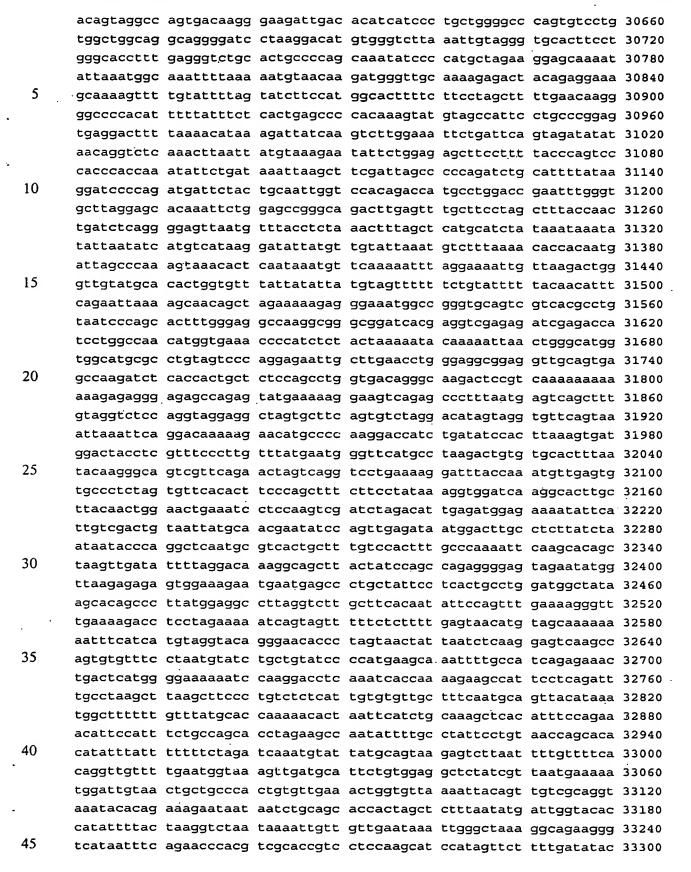
	tatacataag	tattttgaac	taatttcctg	ttttctcaac	cacttgctgt	cttcatgata	22560
						tgaaggaaag	
						attctggtag	
	tccccagtgt	atcaatacat	tatttttctt	tagaaaataa	accaacccaa	ggaaaaatgg	22740
5	tgggcaggtc	ctggtgaata	tggctgtgat	aattatatta	gcaatctctt	tggctaatat	22800
	ttgaagccca	aataattgaa	tcacaatgat	ctctccccag	aaaatatata	aaatgcacct	22860
	tggaatctag	aaggcctttt	agtctgcaaa	agaaaccttc	ttaatcataa	gcagcagaag	22920
	tcccatttac	caaattggaa	agttaaagtt	acaaagcatc	aatcatcaga	cttccattca	22980
	gggatggcaa	ttgggagtaa	gactttttag	taaagaaact	aaacacaaag	tcattagact	23040
10	ctgtaaaagt	cttaccaaat	ttgattctgg	aacacctatt	ctatttccgt	aaagatgatg	23100
	aattcggagc	caaatgttct	tttcatgaag	gatttgaaaa	ctgtccatga	aaataacgca	23160
	atcaaccttt	tagcttgaga	ctctattcac	tgattagatt	tttttaaata	ctgatgggcc	23220
	tgcttctcag	aagtgacaag	gatgggcctc	aatctcaatt	tttgtaatac	atgttccatt	23280
	tgccaatgag	aaatatcagg	ttactaattt	ttcttctatt	tttctagtgc	catttccatg	23340
15	tggaagagtt	tctgtttcac	aaacttctaa	gctcacccgt	gctgaggctg	tttttcctga	23400
	tgtggactat	gtaaattcta	ctgaagctga	aaccattttg	gataacatca	ctcaaagcac	23460
	ccaatcattt	aatgacttca	ctcgggttgt	tggtggagaa	gatgccaaac	caggtcaatt	23520
	cccttggcag	gtactttata	ctgatggtgt	gtcaaaactg	gagctcagct	ggcaagacac	23580
	aggccaggtg	ggagactgag	gctattttac	tagacagacc	tattgggatg	tgagaagtat	23640
20 (	ttaggcaagt	ttcagcacta	accaatgtga	gaaggcctcc	agagatgagc	agttggtgaa	23700
	agagaggctc	aaaaccagct	accatacagg	tcaagaagaa	tttggcatta	aggaaacagc	23760
	atagcaggat	tccagacagg	caactggtca	acaacatgaa	ggtctggaag	aaaggtcgca	23820
	gtactcaggt	tcagggcact	acttcagctt	cagcccttgc	aaaaactggt	gagagttgga	23880
						gactcaagga	
2,5						gtcttgatct	
	accactatag	ttctcgtggt	agcatcagaa	tcacctggga	acgtagaaat,	gcaaattctc	24060
						gcagtctgtg	
						aggccaggtg	
						attgcttaaa	
30			•			ttaaaaaaaa	
						aggaggctga	
						tcacaccact	
						agaaaaacac	
25						ataatgcact	
35						tgcaaacaac	
						tacatttgta	
						tctgtgagaa	
						tccttatata	
40						ttaaattaac	
40						tttaagcaca	
•						atagaacact	
						acttaaaact	
•						tgaggttaga	
45 ·					_	gtgacaattt	
70	aagccagaat	caggtaatgt	cctcagtttt	aacagcattg	gaattttctg	ggactagctg	25200











	ccctattatc	actcatttca	gtgaggtaca	attagttctt	gatgtagcca	tttccatacc	.33360
	agaaggcctt	cccaaaaatc	agtgtcatgt	caccgatcct	tttatctctg	gtgcttggca	33420
	caacctgtag	caggtcctca	gaaaacaaac	atttgaatta	atggccaaat	gagtttgtgc	33480
	tcaaaaaagg	ggtgaggata	cttgaaattt	ggaaaatcta	ggataattca	tgactagtgg	33540
5	attcattatc	accaatgaaa	ggcttataac	agcatgagtg	aacagaacca	tctctatgat	33600
	agtcctgaat	ggctttttgg	tctgaaaaat	atgcattggc	tctcattaca	tttaaccaaa	33660
	attatcacaa	tataagaatg	agatctttaa	cattgccaat	taggtcagtg	gtcccaagta	33720
•	gtcacttaga	aaatctgtgt	atgtgaaata	ctgtttgtga	cttaaaatga	aatttattt	33780
	taataggtga	acataatatt	gaggagacag	aacatacaga	gcaaaagcga	aatgtgattc	33840
10	gaattattcc	tcaccacaac	tacaatgcag	ctattaataa	gtacaaccat	gacattgccc	33900
	ttctggaact	ggacgaaccc	ttagtgctaa	acagctacgt	tacacctatt	tgcattgctg	33960
	acaaggaata	cacgaacatc	ttcctcaaat	ttggatctgg	ctatgtaagt	ggctggggaa	34020
	gagtcttcca	caaagggaga	tcagctttag	ttcttcagta	ccttagagtt	ccacttgttg	34080
	accgagccac	atgtcttcga	tctacaaagt	tcaccatcta	taacaacatg	ttctgtgctg	34140
15	gcttccatga	aggaggtaga	gattcatgtc	aaggagatag	tgggggaccc	catgttactg	34200
	aagtggaagg	gaccagtttc	ttaactggaa	ttattagctg	gggtgaagag	tgtgcaatga	34260
	aaggcaaata	tggaatatat	accaaggtat	cccggtatgt	caactggatt	aaggaaaaaa	34320
						gaaaattaac	
						tatacattct	
20						gcaaattgat	
						tttctaaggg	
						atactatggt	
	tctccactat						
0.5						cagtacagga	
25						cacaggagta	
						ttcctcaatc	
						tecetetete	
						tatacttctg	
20 :						cttgcttttc	
30						tttcagagag	
	•					ataaatatat	
						atggaagcaa	
						ttcacgaagg	
35						attgtactct	
33						tgcaatctac	
						tcagagccaa	
						ttaactagca	
					_	aatcagccaa	
40						agttaatcct	
. *0						tttgtgaatt tgtgaatcag	
						ctgccaagct	
			•			agttccttta	
						tatttttaaa	
45						ggccaaggtg	
		222	222222	Lucuccuy	Jaccoccyga	aaccaaggeg	2000

	ggcggatcac	ctgaggttag	gagtttcagg	ccagcctggc	caacatggtg	aaaccctgtc	36060
	tctactaaaa	atagaacaat	tagctgggct	tggtaatgtg	cacctataat	cccagctact	36120
	ggggaggctg	aggcaggaga	atcacttgag	cctggggagc	agggggtgcg	ggaggttgca	36180
	gtgagacaag	atcgcaccag	tgcactcccc	atcctgggtg	acagagtgag	actctgtctc	36240
5	aaagaaaata	aataaataaa	tacatttctt	gaggcgtttc	ttgttaaatc	attcatggag	36300
	aggcatccca	aacaccacat	tcaacaaaac	actctgaaaa	atgttttcaa	atgcaatata	36360
	acacagcaga	gatttgatgc	tctgttatcc	agttttcata	taaggctgtg	tgagctgtgt	36420
	cccagagagg	acagtggtct	gaatccacct	gagacagaat	tgggtctaac	taactgtgag	36480
	tatggccttc	aataagtcac	tctccatttg	ggaatttgat	ttctccactt	gtataatgag	36540
10	agtatttgac	aggatgctct	cccaaatccc	ttgcaatttt	gttagtctgt	gatttcatgt	36600
	ttttatttt	attccttcat	ccaacaaata	gtcaaggagt	aattgctgtg	tgccaaatac	36660
	caacagtatt	cattaaattg	taattcagat	tttatatata	tataaataat	gtataatgtg	36720
	tataaattgc	tttgtgagtg	cctactacac	tgctagacag	tagttgctca	ataacttgtt	36780
	agctgaatca	gaatccatgt	ttatcccaga	gtagcaatta	gtcttgcatc	gagtatcgtg	36840
15	aaagaaggcc	acacttaaat	aagaataatg	cctggggttt	aggttttatg	aaaaaatgaa	36900
	aggaaattag.	ttctgctttt	gttgactaaa	ggaagggaag	agagaagaga	cactataatt	36960
	gtctgcctca	gatttaagga	ggaggctaat	tcatgcatta	aacacgttac	ttcaaatttg	37020
	aatgaccaaa	ggtctgtagc	ctcagcactt	caaaattggt	aaaagtaaga	cactctggcc	37080
	ttgtttccat	agagaccacc	ccttacaaag	gcaccaatgg	gaaactggcc	tcaggactcc	37140
20	tgttattggt	cttctctgtg	gcagagaaag	gagctcttgg	acccataaat	ctctgagcca	37200
	cagttctttt	tgccatgggc	tcaaaaatga	ttgaattcat	catgagccac	ctgtggcata	37260
	ttgccacact	aaacatgtgg	ggcctttaag	ctcactaaga	gccaatgtct	tcagagccag	37320
	ccctggcttg	attctaccta	gggcatttgc	agttgccata	taagaatcat	tagtgctttc	37380
	aaaattactg	tagatacttt	gcctaaatag	actaaaacat	gctgccgtca	tattggaagt	37440
25	gacagattaa	aatagaactc	ttgccaagtg	aaggaaagtg	tgctaatata	atgcagtcat	37500
	tttaacttgc	tgtttaagtg	tgattgtttt	tagttctttt	gaatattatt	tgttttatac	37560
	tgacaggaac	gaagtactgt	ccaattttct	ctgccaagga	aaaaagaaaa	ggtgttcttc	37620
	cttacttacc	tgaaccaaaa	cagaccagtt	tacaaaattg	cctaattata	attgctaaac	37680
	aagttccgaa	tgcttacagt	ctaatccaag	aatgtcagag	ctgcaagggc	ccttaaacac	37740
30	catccaatcc	actccactca	tttagcagat	gaagagattg	agggcaacat	aaggccaggc	37800
•	ccaagataac	acaatgacag	ccaggactag	agctcaagtc	tcccaccctg	cactttgaaa	37860
	gaataatgct	ttcaactgga	gtacattaac	tctactgtct	atatttttag	ggcagctggg	37920
	gcattctgca	ttggtggcaa	tcctctcaac	aaccctggga	ctgaaaactg	cctggaattc	37980
	ttactaacaa	ttctctaatt	gaccaaaagg	tgacgaaatc	aaggagacca	ataaggtagc	38040
35	cttggaaagc	aagagtggc				•	38059

<210> 5

<211> 29

<212> PRT

40 <213> Homo sapiens

<400> 5

Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1 5 10 15

5 Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr 20 25

<210> 6

<211> 55 ·

10 <212> PRT

<213> Homo sapiens

<400> 6

Val Phe Leu Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys
15 1 5 10 15

Arg Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu
20 25 30

20 Arg Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val
35 40 45

Phe Glu Asn Thr Glu Arg Thr 50 55

25

<210> 7

<211> 8

<212> PRT

<213> Homo sapiens

30

<400> 7

Thr Glu Phe Trp Lys Gln Tyr Val

35 <210> 8

<211> 38

<212> PRT

<213> Homo sapiens

<400> 8

Asp Gly Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys

1 5 10 15

5 Lys Asp Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu
20 25 30

Gly Lys Asn Cys Glu Leu

. 35

10

<210> 9

<211> 43

<212> PRT

<213> Homo sapiens

15

<400> 9

Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe Cys Lys

1 5 10 15

20 Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly Tyr Arg
20 25 30

Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala

25

<210> 10

<211> 68

<212> PRT

<213> Homo sapiens

30

<400> 10

Val Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu

1 5 10 15

Thr Arg Ala Glu Ala Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr 20 25 30

Glu Ala Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe
35 40 45

40

Asn Asp Phe Thr Arg Val Val Gly Glu Asp Ala Lys Pro Gly Gln 50 55 60

Phe Pro Trp Gln

45 65

<210> 11 <211> 38 <212> PRT <213> Homo sapiens <400> 11 Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly Val Lys Ile Thr Val Val Ala <210> 12 <211> 182 <212> PRT <213> Homo sapiens <400> 12 Gly Glu His Asn Ile Glu Glu Thr Glu His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His Glu Gly Gly Arg Asp Ser Cys 

- 19 -



Gln Gly Asp Ser Gly Gly Pro His Val Thr Glu Val Glu Gly Thr Ser 130 135 140

Phe Leu Thr Gly Ile Ile Ser Trp Gly Glu Glu Cys Ala Met Lys Gly

5 145 150 155 160

Lys Tyr Gly Ile Tyr Thr Lys Val Ser Arg Tyr Val Asn Trp Ile Lys
165 170 175

10 Glu Lys Thr Lys Leu Thr 180

<210> 13

<211> 2410

15 <212> DNA

20

25

30

35

40

45

<213> Mus musculus

#### <400> 13

gaattccgtg gatgtgcttt aaaaccacac ctaacgtttg agcacaagtc tcacgaactg 60 gegetacaac tteateagaa acgaagtete caaatetgte caacgeaaaa acacaaagga 120 gtctaatgac taagtcttcc aaccacaact gtctgctgcg cccggaaaac aagccggggc 180 tetggggace eggggeteag geogeetege teeggeetag eecegeeace ttagttgtgt 240 catececegg geatgetgag catecececg eggeteegge acagaegeee ggaeeteagg 300 tetetgeete egegegggg eeeggeeetg tggeeggagg gageggeegg atggagegga 360 ggatgaaagg cggatacttg gaccagcgag tgccctacac cttctgcagc aaatctcccg 420 gaaatgggag cttgggcgaa gcgctgatgg tcccgcaggg aaagctcatg gacccgggct 480 ccctgccgcc ttccgactca gaagatctct tccaggacct cagtcacttc caagagacgt 540 ggctcgcaga agctcaggta ccggacagtg atgagcagtt tgttcctgat ttccattcag 600 aaaacttagc tttccatagc cccaccacca ggatcaagaa ggaaccccag agtccccgca 660 cagaccccgc cctgtcctgc agcaggaagc caccactccc ctaccaccat ggagagcagt 720 gcctttactc cagacaaatc gccatcaagt cccccgctcc cggtgcccct ggacagtcgc 780 ccctgcagcc cttttccagg gcagaacagc agcagagcct cctgagagcc tccagctctt 840 cccagtecca ecetggecae gggtacettg gtgageaeag eteegtette cageageeeg 900 tggacatgtg ccactcettc acateteetc agggaggggg ccgggaacet eteccagece 960 cctatcaaca ccaactgtcg gagcctgcc caccctaccc ccagcagaac ttcaagcagg 1020 agtaccatga ccccctgtac gaacaggctg gccagcccgc ttcaagccag ggtggggtca 1080 gtgggcacag gtacccaggg gcggggtgg tgatcaaaca ggagcgcaca gacttcgcct 1140 acgactcaga tgtccctgga tgtgcatcaa tgtacctcca cccagagggc ttctctggac 1200 cetetecagg tgatggagtg atgggttatg getatgaaaa atccettega ceatteceag 1260 atgatgtctg cattgtccct aaaaaatttg aaggagacat caagcaggaa gggattggag 1320 ctttccggga ggggccaccc taccagcgcc ggggtgcctt acaactgtgg cagtttctgg 1380 tggccctgct ggatgaccca acaaatgctc atttcattgc ttggacaggc cggggaatgg 1440 agtttaaact aattgaacct gaagaggttg ccaggctctg gggtatccag aagaaccggc 1500 cagccatgaa ttatgacaag ctgagccgct cgctgcgata ctattatgag aaaggcatca 1560 tgcagaaggt ggctggcgaa cgctacgtgt acaagtttgt gtgcgagccg gaggcctgt 1620

```
tetetetgge etteceagat aateaaegte eagetetgaa ggetgagttt gaeeggeeag 1680
      tcagtgagga ggacacagtc cctttgtccc acttggatga gagtcctgcc tacctcccag 1740
      aactcactgg ccccgctccg cccttcggcc acagaggtgg atattcttac taggcaccag 1800
      tggcttcccc ttgacatggt ggggttgctc agtgtatata tcaactgatt tggtattggt 1860
 5
      gaaggccctc tttctgatgc ctgtagaagt ctctggggtc agagctccac tatcccatct 1920
      gatactectg gecagactea getgetaace agagtetgeg ggaaagacag tggaggcagg 1980
      ccaaatctaa aggcagtagc tgaagttcgc tgtggctcac ctgtaccttc agttcagctt 2040
      ggcctctgcc taggtcttgc tcagaggcca agttcctcac ccccaccaca gagatccagt 2100
      gttctattct ggggacatac agggacttcc cttgtttatt atggcaacag ggccaagggg 2160
10
      attotoagaa caccotgtgt otoccototo coaaccocco atgggagaca aagttotgco 2220
      tggcttctgc cctgaacagg ggggtcctgt gttcttggtg ctgtgctctg ggaggcagga 2280
      gcatgtgggc ggcagctggg ggggggtgtg gaagtagaga tggctctctg ccctaggcct 2340
      acccaggeet aattecacet ttgeetetta tgeeagacet taataaagee tetgettete 2400
      cccggaattc
                                                                         2410
15
      <210> 14
      <211> 32
      <212> DNA
      <213> Artificial Sequence
20
      <220>
      <223> Description of Artificial Sequence: Synthetic
      <400> 14
25
      taacaggatc cggcctctca ctaactaatc ac
                                                                         32
      <210> 15
      <211> 31
      <212> DNA
30
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
35
      <400> 15
      caactggatc caagattcaa gatagaagga t
                                                                         31
      <210> 16
      <211> 29
40
      <212> DNA
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
45
```

	WO 00/75279	PCT/US00/15728
	<400> 16	
	ggacaggate ceccaaactt tteaggeae	29
	<210> 17	
5	<211> 29	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
10	<223> Description of Artificial Sequence: Synthet	ic
	<400> 17	
	caagcatgca tctagtgtta gtggaagag	29
15	<210> 18	
	<211> 29	
	<212> DNA	
	<213> Artificial Sequence	
20	<220>	
	<223> Description of Artificial Sequence: Synthet	ic .
	<400> 18	
25	caagcatgca aatattaact caaaatgga	29
	<210> 19	·
	<211> 28	
	<212> DNA	
	<213> Artificial Sequence	
30		
	<220>	
	<223> Description of Artificial Sequence: Synthet	cic
	<400> 19	
35	caagcatgct gttgtttttt gttttaac	28
	<210> 20	
	<211> 29	
	<212> DNA	
40	<213> Artificial Sequence	
	<220>	
	<223> Description of Artificial Sequence: Synthet	ic

	WO 00/75279		PCT/US00/15728
	<400> 20		
	caagcatgca gccattcagt cgaggaagg		29
	<210> 21		
5	<211> 27		
	<212> DNA		
	<213> Artificial Sequence		•
	<220>		
10	<223> Description of Artificial Sequence:	Synthetic	
	<400> 21		
	caagcatgcg atcccttcct tatacct		27
15	<210> 22		
	<211> 27		
	<212> DNA		•
	<213> Artificial Sequence		
20	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 22		•
25	taagettaae etttgetage agattgt		27
	<210> 23		
	<211> 29		
	<212> DNA		
	<213> Artificial Sequence		
30			
	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 23 .		
35	cttggtacca gecattcagt cgaggaagg		29
	<210> 24		
	<211> 29		
	<212> DNA		
40	<213> Artificial Sequence		
	<220>		
	<223> Description of Artificial Sequence:	Synthetic	•

	WO 00/75279		PCT/US00/15728
	<400> 24		
	cttggtacca tatgaatcct ttcatagat		29
•	<210> 25		
5	<211> 23		
	<212> DNA		
	<213> Artificial Sequence		
	<220>		
10	<223> Description of Artificial Sequence:	Synthetic	
	<400> 25		
	ctgtgggaac acacagattt tgg	•	23
15	<210> 26		
	<211> 19		
	<212> DNA		•
	<213> Artificial Sequence		
20	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 26 .		
25	ggaataattc gaatcacat		19
	<210> 27		
	<211> 20		
	<212> DNA		
	<213> Artificial Sequence	•	
30			
	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 27		
35	ccaatctgct cacacaggat		20
	<210> 28		
	<211> 20		
	<212> DNA		•
40	<213> Artificial Sequence		
	<220>		
	<223> Description of Artificial Sequence:	Synthetic	

### WO 00/75279 PCT/US00/15728 <400> 28 ccttgaggct gtccaagtga . 20 <210> 29 5. <211> 20 <212> DNA <213> Artificial Sequence <220> 10 <223> Description of Artificial Sequence: Synthetic <400> 29 atggttaact gacttacgaa 20 15 <210> 30 <211> 26 <212> DNA <213> Artificial Sequence 20 <220> <223> Description of Artificial Sequence: Synthetic <400> 30 gctccatttt gagttaatat ttgtgt 26 25 <210> 31 <211> 22 <212> DNA <213> Artificial Sequence 30 <220> <223> Description of Artificial Sequence: Synthetic <400> 31 35 ttcagtcgag gaaggatagg gt 22 <210> 32 <211> 22 <212> DNA 40 <213> Artificial Sequence <220> . <223> Description of Artificial Sequence: Synthetic

	W U 00//52	2.79			PC1/USUU/15/2
	<400> 32				
•	ttcagtcggt	tġgtgatagg	gt	•	22
_	<210> 33				
5	<211> 11				
	<212> DNA				
	<213> Homo	sapiens			
	<400> 33				
10	tcgaggaạgg	a			11
	<210> 34	•			
	<211> 15	•			
	<212> DNA				
15	<213> Homo	sapiens			
	400 04				
	<400> 34				
	agtcgaggaa	ggata			15
20	<210> 35				
	<211> 19				•
	<212> DNA				
	<213> Homo	sapiens			
25	<400> 35				
	tcagtcgagg	aaggatagg.			19
	<210> 36				
	<211> 23				•
30	<212> DNA				
	<213> Homo	sapiens		•	
		_			
	<400> 36				
	attcagtcga	ggaaggatag	ggt		23
35					
	<210> 37				
	<211> 27				
	<212> DNA				
40 .	<213> Homo	sapiens			
40	400- 37				
	<400> 37		2000100		2.7
	CCarreage	gaggaaggat	ayyytyg		27

WO 00/75279

<400> 42

	gaattccagg	ttggaggggc	ggcaacctcc	tgccagcctt	caggccactc	tcctgtgcct	60
	gccagaagag	acagagcttg	aggagagctt	gaggagagca	ggaaaggtgg	aacattgctg	120
	ctgctgctca	ctcagttcca	caggtgggag	gaacagcagg	gcttagagtg	ggggtcattg	180
5	tgcagatggg	aaaacaaagg	cccagagagg	ggaagaaatg	cctaggagct	accgagggca	240
	ggcgacctca	accacagccc	agtgctggag	ctgtgagtgg	atgtagagca	gcggaatatc	300
	cattcagcca	gctcagggga	aggacagggg	ccctgaagcc	aggggatgga	gctgcaggga	360
	agggagctca	gagagaaggg	gagggagtc	tgagctcagt	ttcccgctgc	ctgaaaggag.	420
	ggtggtacct	actcccttca	cagggtaact	gaatigagaga	ctgcctggag	gaaagctctt	480
10	caagtgtggc	ccacccacc	ccagtgacac	cagcccctga	cacgggggag	ggagggcagc	540
•	atcaggaggg	gctttctggg	cacacccagt	acccgtctct	gagctttcct	tgaactgttg	600
	cattttaatc	ctcacagcag	ctcaacaagg	tacataccgt	caccatcccc	attttacaga	660
	tagggaaatt	gaggctcgga	gcggttaaac	aactcacctg	aggcctcaca	gccagtaagt	720
	gggttccctg	gtctgaatgt	gtgtgctgga	ggatcctgtg	ggtcactcgc	ctggtagagc	780
15	cccaaggtgg	aggcataaat	gggactggtg	aatgacagaa	ggggcaaaaa	tgcactcatc	840
	cattcactct	gcaagtatct	acggcacgta	cgccagctcc	caagcaggtt	tgcgggttgc	900
	acagcggagc	gatgcaatct	gatttaggct	tttaaaggat	tgcaatcaag	tgggacccac	960
	tagcctcaac	cctgtacctc	ccctcccctc	cacccccagc	agtctccaaa	ggcctccaac	1020
	aaccccagag	tgggggccat	gtatccaaag	aaactccaag	ctgtatacgg	atcacactgg	1080
20	ttttccagga	gcaaaaacag	aaacagcctg	aggctggtca	aaattgaacc	tcctcctgct	1140
	ctgagcagcc	tagggggcag	actaagcaga	gggctgtgca	gacccacata	aagagcctac	1200
	tgtgtgccag	gcacttcacc	cgaggcactt	cacaagcatg	cttgggaatg	aaacttccaa	1260
	ctctttggga	tgcaggtgaa	acagttcctg	gttcagagag	gtgaagcggc	ctgcctgagg	1320
	cagcacagct	cttctttaca	gatgtgcttc	cccacctcta	ccctgtctca	cggcccccca	1380
25	tgccagcctg	acggttgtgt	ctgcctcagt	catgctccat	ttttccatcg	ggaccatcaa	1440
	gagggtgttt	gtgtctaagg	ctgactgggt	aactttggat	gagcggtctc	tccgctccga	1500
٠	gcctgtttcc	tcatctgtca	aacgggetet	aacccactct	gatctcccag	ggcggcagta	1560
	agtcttcagc	atcaggcatt	ttggggtgac	tcagtaaatg	gtagatcttg	ctaccagtgg	1620
	aacagccact	aaggattctg	cagtgagagc	agagggccag	ctaagtggta	ctctcccaga	1680
30	gactgtctga	ctcacgccac	ccctccacc	ttggacacag	gacgctgtgg	tttctgagcc	1740
	aggtacaatg	actcctttcg	gtaagtgcag	tggaagctgt	acactgccca	ggcaaagcgt	1800
•	ccgggcagcg	taggcgggcg	actcagatcc	cagccagtgg	acttagcccc	tgtttgctcc	1860
	tccgataact	ggggtgacct	tggttaatat	tcaccagcag	cctccccgt	tgcccctctg	1920
	gatccactgc	ttaaatacgg	acgaggacag	ggccctgtct	cctcagcttc	aggcaccacc	1980
35						tggagaggct	
						gttgtggagt	
						tcctattttc	•
						tececeteg	
40						gtctaggtct	
40						caacaaagcg	
						tctgttgtct	
						cttctataga	
			•		•	ctccatcccc	
45						atgggcctcc	
45	ccatgcaccg	tggacctctg	ggccctcctg	tcccagtgga	ggacaggaag	ctgtgagggg	2640



	cactgtcacc	cagggctcaa	gctggcattc	ctgaataatc	gctctgcacc	aggccacggc	2700
	taagctcagt	gcgtgattaa	gcctcataac	cctccaaggc	agttactagt	gtgattccca	2760
	ttttacagat	gaggaagatg	gggacagaga	ggtgaataac.	tggccccaaa	tcacacacca	2820
	tccataattc	gggctcaggc	acctggctcc	agtccccaaa	ctcttgaacc	tggccctagt	2880
5	gtcactgttt	ctcttgggtc	tcaggcgctg	gatggggaac	aggaaacctg	ggctgaactt	2940
	gaggcctctc	tgatgctcgg	tgacttcaga	cagttgctca	acctctctgt	tctcttgggc	3000
	aaaacatgat	aacctttgac	ttctgtcccc	tcccctcacc	ccacccgacc	ttgatctctg	3060
	aagtgttgga	aggatttaat	ttttcctgca	ctgagttttg	gagacaggtc	aaaaagatga	3120
	ccaaggccaa	ggtggccagt	ttcctataga	acgcctctaa	aagacctgca	gcaatagcag	3180
10	caagaactgg	tattctcgag	aacttgctgc	gcagcaggca	cttcttggca	ttttatgtgt	3240
	atttaatttc	acaatagctc	tatgacaaag	tccacctttc	tcatctccag	gaaactgagg	3300
	ttcagagagg	ttaagtaact	tgtccaaggt	cacacagcta	atagcaagtt	gacgtggagc	3360
	aatctggcct	cagagccttt	aattttagcc	acagactgat	gctcccctct	tcatttagcc	3420
	aggctgcctc	tgaagttttc	tgattcaaga	cttctggctt	cagctttgta	cacagagatg	3480
15	attcaatgtc	aggttttgga	gcgaaatctg	tttaatccca	gacaaaacat	ttaggattac	3540
	atctcagttt	tgtaagcaag	tagctctgtg	atttttagtg	agttatttaa	tgctctttgg	3600
	ggctcaattt	ttctatctat	aaaatagggc	taataatttg	caccttatag	ggtaagcttt	3660
	gaggacagat	tagatgatac	ggtgcctgta	aaacaccagg	tgttagtaag	tgtggcaatg	3720
	atggtgacgc	tgaggctgtg	tttgcttagc	atagggttag	gcagctggca	ggcagtaaac	3780
20	agttggataa	tttaatggaa	aatttgccaa	actcagatgc	tgttcactgc	tgagcaggag	3840
	ccccttcctg	ctgaaatggt	cctggggagt	gcagcaggct	ctccgggaag	aaatctacca	3900
	tctctcgggc	aggagctcaa	cctgtgtgca	ggtacaggga	gggcttcctc	acctggtgcc	3960
	cactcatgca	ttacgtcagt	tattcctcat	ccctgtccaa	aggattcttt	tctccattgt	4020
	acagctatga	agctagtgct	caaagaagtg	aagtcattta	ccccaggccc	cctgccagta	4080
25	agtgacaggg	cctggtcaca	cttgggttta	tttattgccc	agttcaacag	gttgtttgac	4140
	cataggcgag	attctcttcc	ctgcaccctg	ccgggttgct	cttggtccct	tattttatgc	4200
	tcctgggtag	aaatggtgcg	agattaggca	gggagtggac	gcttccctgt	ccctggcccc	4260
	gcaaagagtg	ctcccacctg	ccccgatccc	agaaatgtca	ccatgaagcc	ttcattcttt	4320
	tggtttaaag	cttggcctca	gtgtccgtac	accatggggt	ccttggccag	atggcgactt	4380
30	tctcctctcc	agtcgccctc	ccaggcacta	gcttttagga	gtgcagggtg	ctgcctctga	4440
	tagaagggcc	aggagagagc	aggttttgga	gacctgatgt	tataaggaac	agcttgggag	4500
	gcataatgaa	cccaacatga	tgcttgagac	caatgtcaca	gcccaattct	gacattcatc	4560
			gctgtctcag			_	
2.5	cttgttccag	ctttgtcact	gacttgctgt	atagcctcaa	caaggccctg	accetetetg	4680
35			gaaaggagga				
		•	tgtgctcctc		_	_	
	•		ttgagcccag				
			tccatctgtt				
40			caggggctct				
40	•		caggagttca				
			aattagccag			· · · · · · · · · · · · · · · · · · ·	
			agaatttctt	•			
			caccctggcg				
15			acagagggcc				
45	cacaactggg	gagagtgagg	ccccaggacc	agaggattct	tgctaaaggc	caagtggata	5340



	gtgatggccc	tgccaggcta	gaagccacaa	cctctggccc	tgaggccact	cagcatattt	5400
	agtgtcccca	ccctgcagag	gcccaactcc	ctcctgacca	ctgagccctg	taatgatggg	5460
	ggaatttcca	taagccatga	aggactgcac	aaagttcagt	tgggagtgaa	agagaaatta	5520
	aagggagatg	gaaatataca	gcactaattt	tagcaccgtc	ttcagttcta	acaacactag	5580
5	ctagctgaag	aaaatacaaa	catgtattat	gtaatgtgtg	gtctgttcca	tttggattac	5640
	ttagaggcac	gagggccaag	gagaaaggtg	gtggagagaa	accagctttg	cacttcattt	5700
	gttgctttat	tggaaggaaa	cttttaaaag	tccaaggggg	ttgaagaatc	tcaatatttg	5760
	ttatttccag	cttttttct	ccagtttttc	atttcccaaa	ttcaaggaca	cctttttctt	5820
	tgtattttgt	taagatgatg	gttttggttt	tgtgactagt	agttaacaat	gtggctgccg	5880
10	ggcatattct	cctcagctag	gacctcagtt	ttcccatctg	tgaagacggc	aggttctacc	5940
	tagggggctg	caggcaggtg	gtccgaagcc	tgggcatatc	tggagtagaa	ggatcactgt	6000
	ggggcagggc	aggttctgtg	ttgctgtgga	tgacgttgac	tttgaccatt	gctcggcaga	6060
	gcctgctctc	gctggttcag	ccacaggccc	caccactccc	tattgtctca	gccccgggta	6120
	tgaaacatgt	attcctcact	ggcctatcac	ctgaagcctt	tgaatttgca	acacctgcca	6180
15	acccctccct	caaaagagtt	gccctctcta	gatccttttg	atgtaaggtt	tggtgttgag	6240
	acttatttca	ctaaattctc	atacataaac	atcactttat	gtatgaggca	aaatgaggac	6300
	cagggagatg	aatgacttgt	cctggctcat	acacctggaa	agtgacagag	tcagattaga	6360
	tcctaggtct	atctgaagtt	aaaagaggtg	tcttttcact	tcccacctcc	tccatctact	6420
	ttaaagcagc	acaaacccct	gctttcaagg	agagatgagc	gtctctaaag	cccctgacag	6480
20	caagagccca	gaactgggac	accattagtg	acccagacgg	caggtaagct	gactgcagga	6540
	gcatcagcct	attcttgtgt	ctgggaccac	agagcattgt	ggggacagcc	ccgtctcttg	6600
	ggaaaaaaac	cctaagggct	gaggatcctt	gtgagtgttg	ggtgggaaca	gctcccagga	6660
	ggtttaatca	cagcccctcc	atgctctcta	gctgttgcca	ttgtgcaaga	tgcatttccc	6720
	ttctgtgcag	cagtttccct	ggccactaaa	tagtgggatt	agatagaagc	cctccaaggg	6780
25	ctccagcttġ	acatgattct	tgattctgat	ctgacccgat	tctgataatc	gtgggcaggc	6840
	ccattcctct	tcttgtgcct	cattttcttc	ttttgtaaaa	caatggctgt	accatttgca	6900
	tcttagggtc	attgcagatg	aaagtgttgc	tgtccagagc	ctgggtgcag	gacctagatg	6960
	taggattctg	gttctgctac	ttcctcagtg	acattġaata	gctgacctaa	tctctctggc	7020
	tttggtttct	tcatctgtaa	aagaaggata	ttagcattag	cacctcacgg	gattgttaca	7080
30	agaaagcaat	gaattaacac	atgtgagcac	ggagaacagt	gcttggcata	tggtaagcac	7140
	tacgtacatt	ttgctattct	tctgattctt	tcagtgttac	tgatgtcggc	aagtacttgg	7200
	cacaggctgg	tttaataatc	cctaggcact	ttcacgtggt	gtcaatccct	gatcactggg	7260
	agtcatcatg	tgccttgact	cgggcctggc	cccccatct	ctgtcttgca	ggacaatgcc	7320
		tcgtggggca					
35		gatccccagg		•			
		accttcaaca					
		gcacaccagt					
		gcaatgctct					
		ttcaacctca			:		
40		accctcaacc		·			
		gagggcctga					
		gccttcactg			•		
		gagaagggta					
4.5		tttgctctgg					
45	ctgagctgtt	tcccatagaa	acaagcaaaa	atatttctca	aaccatcagt	tcttgaactc	8040

	tccttggcaa	tgcattatgg	gccatagcaa	tgcttttcag	cgtggattct	tcagttttct	8100
						tagattaaac	
	tgtcaaaact	actgacgctc	tgcagaactt	ttcagagcct	ttaatgtcct	tgtgtatact	8220
	gtatatgtag	aatatataat	gcttagaact	atagaacaaa	ttgtaataca	ctgcataaag	8280
5	ggatagtttc	atggaacata	ctttacacga	ctctagtgtc	ccagaatcag	tatcagtttt	8340
	gcaatctgaa	agacctgggt	tcaaatcctg	cctctaacac	aattagcttt	tgacaaaaac	8400
	aatgcattct	acctctttga	ggtgctaatt	tctcatctta	gcatggacaa	aataccattc	8460
	ttgctgtcag	gtttttttag	gattaaacaa	atgacaaaga	ctgtggggat	ggtgtgtggc	8520
	atacagcagg	tgatggactc	ttctgtatct	caggctgcct	tcctgcccct	gaggggttaa	8580
10	aatgccaggg	tcctgggggc	cccagggcat	tctaagccag	ctcccactgt	cccaggaaaa	8640
	cagcataggg	gagggaggt	gggaggcaag	gccaggggct	gcttcctcca	ctctgaggct	8700
	cccttgctct	tgaggcaaag	gagggcagtg	gaggcaagcc	aggctgcagt	cagcacagct	8760
	aaagtcctgg	ctctgctgtg	gccttagtgg	gggcccaggt	ccctctccag	ccccagtctc	8820
	ctccttctgt	ccaatgagaa	agctgggatc	aggggtccct	gaggcccctg	tccactctgc	8880
15	atgcctcgat	ggtgaagctc	tgttggtatg	gcagagggga	ggctgctcag	gcatctgcat	8940
	ttcccctgcc	aatctagagg	atgaggaaag	ctctcaggaa	tagtaagcag	aatgtttgcc	9000
	ctggatgaat	aactgagctg	ccaat.taaca	aggggcaggg	agccttagac	agaaggtacc	9060
	aaatatgcct	gatgctccaa	cattttattt	gtaatatcca	agacaccctc	aaataaacat	9120
	atgattccaa	taaaaatgca	cagccacgat	ggcatctctt	agcctgacat	cgccacgatg	9180
20	tagaaattct	gcatcttcct	ctagttttga	attatcccca	cacaatcttt	ttcggcagct	9240
	tggatggtca	gtttcagcac	cttttacaga	tgatgaagct	gagcctcgag	ggatgtgtgt	9300
	cgtcaagggg	gctcagggct	tctcagggag	gggactcatg	gtttcttatt	ctgctacact	9360
	cttccaaacc	ttcactcacc	cctggtgatg	cccaccttcc	cctctctcca	ggcaaatggg	9420
	agagaccctt	tgaagtcaag	gacaccgagg	aagaggactt	ccacgtggac	caggtgacca	9480
25	ccgtgaaggt	gcctatgatg	aagcgtttag	gcatgtttaa	catccagcac	tgtaagaagc	9540
	tgtccagctg	ggtgctgctg	atgaaatacc	tgggcaatgc	caccgccatc	ttetteetge	9600
	ctgatgaggg	gaaactacag	cacctggtaa	atgaactcac	ccacgatatc	atcaccaagt	9660
						aagctgccca	
20						tgaggagggt	
30	gaaccccaca	tccctgggtc	actgctactc	tgtataaact	tggcttccag	aatgaggcca	9840
						agggtgagga	
						cttcccaata	
,						caaaaaggaa	
25				•		gaatgtccct	
35						ggaattggaa	
						tgctgttaga	
						atgtgtgtgc	
		_			_	aaaaccaaga	
40					•	ctctgttcca	
. 40						ggctgagagg	
						tcagggtggc	
						ctcttctctt	
-						gggtgtaggg	
15						agatgagaca	
45	ggcttccagg	acacctagaa	tatcagagga	ggtggcattt	caagcttttg	tgattcattc	10740



```
gatgttaaca ttctttgact caatgtagaa gagctaaaag tagaacaaac caaagccgag 10800
      ttcccatctt agtgtgggtg gaggacacag gagtaagtgg cagaaataat cagaaaagaa 10860
      aacacttgca ctgtggtggg tcccagaaga acaagaggaa tgctgtgcca tqccttqaat 10920
      ttettttetg caegacaggt etgecagett acatttacec aaactgteea ttactggaac 10980
 5
      ctatgatctg aagagcgtcc tgggtcaact gggcatcact aaggtcttca gcaatggggc 11040
      tgacctctcc ggggtcacag aggaggcacc cctgaagctc tccaaggtga gatcaccctg 11100
      acgacettgt tgcaccatgg tatetgtagg gaagaatgtg tgggggetge agcactgtee 11160
      tgaggctgag gaaggggccg agggaaacaa atgaagaccc aggctgagct cctgaagatg 11220
      cccgtgattc actgacacyg gacggtgggc aaacagcaaa gccaggcagg ggctgctgtg 11280
10
      cagctggcac tttcggggcc tcccttgagg ttgtgtcact gaccctgaat ttcaactttg 11340
      cccaagacct tctagacatt gggccttgat ttatccatac tgacacagaa aggtttgggc 11400
      taagttgttt caaaggaatt tetgaeteet tegatetgtg agatttggtg tetgaattaa 11460
      tgaatgattt cagctaaagt gacacttatt ttggaaaact aaaggcgacc aatgaacaac 11520
      ctgcagttcc atgaatggct gcattatctt ggggtctggg cactgtgaag gtcactgcca 11580
15
      gggtccgtgt cctcaaggag cttcaagccg tgtactagaa aggagagagc cctggaggca 11640
      gacgtggagt gacgatgctc ttccctgttc tgagttgtgg gtgcacctga gcagggggag 11700
      aggegettgt caggaagatg gacagagggg agceageeee ateageeaaa geettgagga 11760
      ggagcaaggc ctatgtgaca gggagggaga ggatgtgcag ggccagggcc gtccaggggg 11820
      agtgagcgct tectgggagg tgtecacgtg ageettgete gaggeetggg ateageetta 11880
20
      caacgtgtct ctgcttctct cccctccagg ccgtgcataa ggctgtgctg accatcgacg 11940
      agaaagggac tgaagctgct ggggccatgt ttttagaggc catacccatg tctatccccc 12000
      ccgaggtcaa gttcaacaaa ccctttgtct tcttaatgat tgaacaaaat.accaagtctc 12060
      ccctcttcat gggaaaagtg gtgaatccca cccaaaaata actgcctctc gctcctcaac 12120
      ccctccctc catccctggc cccctccctg gatgacatta aagaagggtt gagctggtcc 12180
25
      ctgcctgcat gtgatctgta aatccctggg atgttttctc tg
                                                                         12222
      <210> 43
      <211> 1599
```

<212> DNA

30 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

35 <400> 43

40

45

caccagcatc attractica attractora ctactetge catgaagata atagtittea 60 ggcggattge ctcagateac actateteca cttgeceage cetgtggaag attageggee 120 atgtateca atgtgatagg aactgtaace tetggaaaaa ggaaggttta tettttgtee 180 ttgetgetea ttggettetg ggactgegtg acctgteaeg ggageeetgt ggacatetge 240 acageeaage egegggacat teccatgaat eccatgtgea tttacegete eeggagaag 300 aaggeaactg aggatgaggg etcagaacag aagateeegg aggeeaceaa eeggegtgte 360 tgggaactgt ecaaggeeaa tteeegettt getaceaett tetateagea eetggeagat 420 tecaagaatg acaatgataa eatttteetg teaeeeetga gtateteeae ggettttget 480 atgaceage tgggtgeetg taatgacaee etceageaae tgatggaggt atttaagttt 540 gacaccatat etgagaaaac atetgateag ateeaettet tetttgeeaa actgaactge 600

	cgactctatc	gaaaagccaa	caaatcctcc	aagttagtat	cagccaatcg	cctttttgga	660
	gacaaatccc	ttaccttcaa	tgagacctac	caggacatca	gtgagttggt	atatggagcc	720
	aagctccagc	ccctggactt	caaggaaaat	gcagagcaat	ccagagcggc	catcaacaaa	780
	tgggtgtcca	ataagaccga	aggccgaatc	accgatgtca	ttccctcgga	agccatcaat	840
5	gagctcactg	ttctggtgct	ggttaacacc	atttacttca	agggcctgtg	gaagtcaaag	900
	ttcagccctg	agaacacaag	gaaggaactg	ttctacaagg	ctgatggaga	gtcgtgttca	960
	gcatctatga	tgtaccagga	aggcaagttc	cgttatcggc	gcgtggctga	aggcacccag	1020
	gtgcttgagt	tgcccttcaa	aggtgatgac	atcaccatgg	tcctcatctt	gcccaagcct	1080
	gagaagagcc	tggccaaggt	ggagaaggaa	ctcaccccag	aggtgctgca	ggagtggctg	1140
10	gatgaattgg	aggagatgat	gctggtggtt	cacatgcccc	gcttccgcat	tgaggacggc	1200
	ttcagtttga	aggagcagct	gcaagacatg	ggccttgtcg	atctgttcag	ccctgaaaag	1260
	tccaaactcc	caggtattgt	tgcagaaggc	cgagatgacc	tctatgtctc	agatgcattc	1320
	cataaggcat	ttcttgaggt	aaatgaagaa	ggcagtgaag	cagctgcaag	taccgctgtt	1380
	gtgattgctg	gccgttcgct	aaaccccaac	agggtgactt	tcaaggccaa	caggcccttc	1440
15	ctggttttta	taagagaagt	tcctctgaac	actattatct	tcatgggcag	agtagccaac	1500
	ccttgtgtta	agtaaaatgt	tcttattctt	tgcacctctt	cctatttttg	gtttgtgaac	1560
	agaagtaaaa	ataaatacaa	actacttcca	tctcacatt			1599
					•		
	<210> 44						
20	<211> 23						
	<212> DNA						
	<213> T7 vi	rus					
25	<400> 44						
25	taatacgact	cactataggg	cga				23
	<210> 45		•				
	<211> 24						
	<211> 24 <212> DNA						
30	<213> T3 vi	TILE.					
	1227 23 12	.145					
	<400> 45						
	ttattaaccc	tcactaaagg	qaaq	•			24
35	<210> 46						
	<211> 23			•			
	<212> DNA	•				•	
	<213> SP6 v	rirus					•
40	<400> 46	•					
	atttaggtga	cactatagaa	tac				22

	<21:	0 > 4' 1 > 5! 2 > PI	55 RT		1											-
5.	<21.	3> Mi	is mi	iscu.	lus		•									
	Met	0> 4° Thr		Ser		Asn	His	Asn	Cys		Leu	Arg	Pro	Glu		Lys
	1				5					10					15	
10	Pro	Gly	Leu	Trp 20	Gly	Pro	Gly	Ala	Gln 25	Ala	Ala	Ser	Leu	Arg 30	Pro	Ser
15	Pro	Ala	Thr 35	Leu	Val	Val	Ser	Ser 40	Pro	Gly	His	Ala	Glu 45	His	Pro	Pro
13	Ala	Ala 50	Pro	Ala	Gln	Thr	Pro 55	Gly	Pro	Gln	Val	Ser 60	Ala	Ser	Ala	Arg
20	Gly 65	Pro	Gly	Pro	Val	Ala 70	Gly	Gly	Ser	Gly	Arg 75	Met	Glu	Arg	Arg	Met 80
	Lys	Gly	Gly	Tyr	Leu 85	Asp	Gln	Arg	Val	Pro 90	Tyr	Thr	Phe	Суз	Ser 95	Lys
25	Ser	Pro	Gly	Asn 100	Gly	Ser	Leu	Gly	Glu 105	Ala	Leu	Met	Val	Pro 110	Gln	Gly
30	Lys	Leu	Met 115	Asp	Pro	Gly	Ser	Leu 120	Pro	Pro	Ser	Asp	Ser 125	Glu	Asp	Leu
30	Phe	Gln 130	Asp	Leu	Ser	His	Phe		Glu	Thr	Trp	Leu 140	Ala ·.	Glu	Ala	Gln
35	Val 145	Pro	Asp	Ser	Asp	Glu 150	Gln	Phe	Val	Pro	Asp 155	Phe	His	Ser	Glu	Asn 160
	Leu	Ala	Phe	His	Ser 165	Pro	Thr	Thr	Arg	Ile 170	Lys	Lys	Glu	Pro	Gln 175	Ser
40	Pro	Arg	Thr	Asp 180	Pro	Ala	Leu	Ser	Cys 185	Ser	Arg	Lys	Pro	Pro 190	Leu	Pro
45	Tyr	His	His 195	Gly	Glu	Gln	Cys	Leu 200	Tyr	Ser	Arg	Gln	Ile 205	Ala	Ile	Lys

Ser Pro Ala Pro Gly Ala Pro Gly Gln Ser Pro Leu Gln Pro Phe Ser Arg Ala Glu Gln Gln Ser Leu Leu Arg Ala Ser Ser Ser Gln Ser His Pro Gly His Gly Tyr Leu Gly Glu His Ser Ser Val Phe Gln Gln Pro Val Asp Met Cys His Ser Phe Thr Ser Pro Gln Gly Gly Gly Arg Glu Pro Leu Pro Ala Pro Tyr Gln His Gln Leu Ser Glu Pro Cys Pro Pro Tyr Pro Gln Gln Asn Phe Lys Gln Glu Tyr His Asp Pro Leu Tyr Glu Gln Ala Gly Gln Pro Ala Ser Ser Gln Gly Gly Val Ser Gly His Arg Tyr Pro Gly Ala Gly Val Val Ile Lys Gln Glu Arg Thr Asp . 330 Phe Ala Tyr Asp Ser Asp Val Pro Gly Cys Ala Ser Met Tyr Leu His Pro Glu Gly Phe Ser Gly Pro Ser Pro Gly Asp Gly Val Met Gly Tyr Gly Tyr Glu Lys Ser Leu Arg Pro Phe Pro Asp Asp Val Cys Ile Val Pro Lys Iys Phe Glu Gly Asp Ile Lys Gln Glu Gly Ile Gly Ala Phe Arg Glu Gly Pro Pro Tyr Gln Arg Arg Gly Ala Leu Gln Leu Trp Gln . 405 Phe Leu Val Ala Leu Leu Asp Asp Pro Thr Asn Ala His Phe Ile Ala Trp Thr Gly Arg Gly Met Glu Phe Lys Leu Ile Glu Pro Glu Glu Val 

	Ala Arg Leu Trp Gly Ile Gln Lys Asn Arg Pro Ala Met Asn Tyr Asp 450 455 460	
5	Lys Leu Ser Arg Ser Leu Arg Tyr Tyr Tyr Glu Lys Gly Ile Met Gln 465 470 475 480	
	Lys Val Ala Gly Glu Arg Tyr Val Tyr Lys Phe Val Cys Glu Pro Glu 485 490 495	
10	Ala Leu Phe Ser Leu Ala Phe Pro Asp Asn Gln Arg Pro Ala Leu Lys 500 505 510	
15	Ala Glu Phe Asp Arg Pro Val Ser Glu Glu Asp Thr Val Pro Leu Ser 515 520 525	
13	His Leu Asp Glu Ser Pro Ala Tyr Leu Pro Glu Leu Thr Gly Pro Ala 530 535 540	
20	Pro Pro Phe Gly His Arg Gly Gly Tyr Ser Tyr 545 550 555	
25	<210> 48 <211> 8 <212> DNA <213> Artificial Sequence	
30	<220> <223> Description of Artificial Sequence: Synthetic <400> 48	
. 35	cccggaag <210> 49 <211> 1843	8
	<212> DNA <213> Homo sapiens	
40	<pre>&lt;400&gt; 49  ctgcaggggg ggggggggg gggggctgtc atggcggcag gacggcgaac ttgcagtatc tccacgaccc gcccctacag gtgccagtgc ctccagaatg tggcagctca caagcctcct gctgttcgtg gccacctggg gaatttccgg cacaccagct cctcttgact cagtgttctc cagcagcgag cgtgcccacc aggtgctgcg gatccgcaaa cgtgccaact ccttcctgga ggagctccgt cacagcagcc tggagcgga gtgcatagag gagatctgtg acttcgagga</pre>	120 180 240

ggccaaggaa attttccaaa atgtggatga cacactggcc ttctggtcca agcacgtcga 360

45



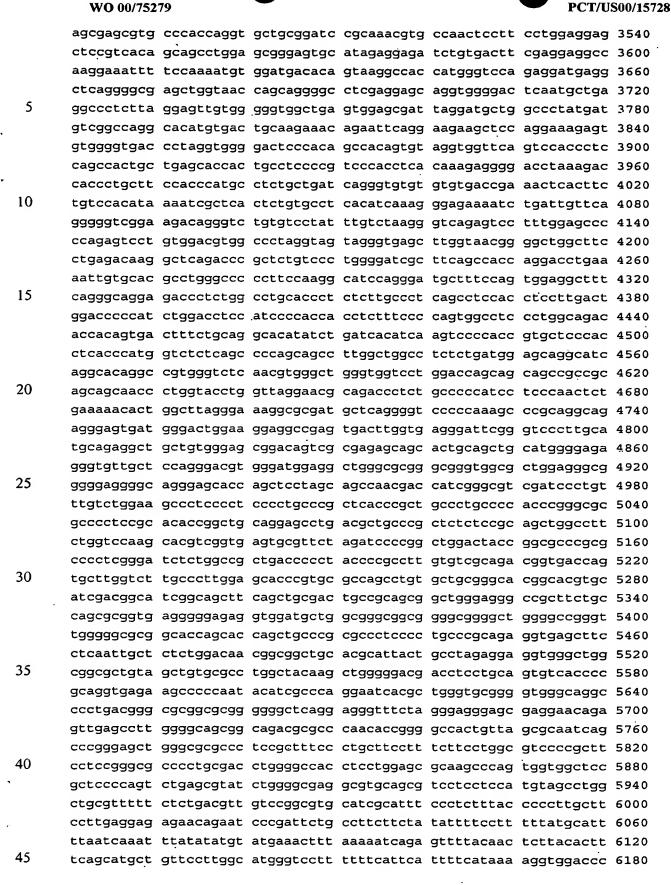
	cggtgaccag	tgcttggtct	tgcccttgga	gcacccgtgc	gccagcctgt	gctgcgggcá	420
	cggcacgtgc	atcgacggca	tcggcagctt	cagctgcgac	tgccgcagcg	gctgggaggg	480
	ccgcttctgc	cagcgcgagg	tgagcttcct	caattgctcg	ctggacaacg	gcggctgcac	540
	gcattactgc	ctagaggagg	tgggctggcg	gcgctgtagc	tgtgcgcctg	gctacaagct	60'0
5	gggggacgac	ctcctgcagt	gtcaccccgc	agtgaagttc	ccttgtggga	ggccctggaa	660
	gcggatggag	aagaagcgca	gtcacctgaa	acgagacaca	gaagaccaag	aagaccaagt	720
	agatccgcgg	ctcattgatg	ggaagatgac	caggcgggga	gacagcccct	ggcaggtggt	780
	cctgctggac	tcaaagaaga	agctggcctg	cggggcagtg	ctcatccacc	cctcctgggt	840
	gctgacagcg	gcccactgca	tggatgagtc	caagaagctc	cttgtcaggc	ttggagagta	900
10	tgacctgcgg	cgctgggaga	agtgggagct	ggacctggac	atcaaggagg	tcttcgtcca	960
	ccccaactac	agcaagagca	ccaccgacaa	tgacatcgca	ctgctgcacc	tggcccagcc	1020
	cgccaccctc	tcgcagacca	tagtgcccat	ctgcctcccg	gacagcggcc	ttgcagagcg	1080
	cgagctcaat	caggccggcc	aggagaccct	cgtgacgggc	tggggctacc	acagcagccg	1140
	agagaaggag	gccaagagaa	accgcacctt	cgtcctcaac	ttcatcaaga	ttcccgtggt	1200
15	cccgcacaat	gagtgcagcg	aggtcatgag	caacatggtg	tctgagaaca	tgctgtgtgc	1260
	gggcatcctc	ggggaccggc	aggatgcctg	cgagggcgac	agtggggggc	ccatggtcgc	1320
	ctccttccac	ggcacctggt	tcctggtggg	cctggtgagc	tggggtgagg	gctgtgggct	1380
	ccttcacaac	tacggcgttt	acaccaaagt	cagccgctac	ctcgactgga	tccatgggca	1440
	catcagagac	aaggaagccc	cccagaagag	ctgggcacct	tagcgaccct	ccctgcaggg	1500
20	ctgggctttt	gcatggcaat	ggatgggaca	ttaaagggac	atgtaacaag	cacaccggcc	1560
	tgctgttctg	tccttccatc	cctcttttgg	gctcttctgg	agggaagtaa	catttactga	1620
-	gcacctgttg	tatgtcacat	gccttatgaa	tagaatctta	actcctagag	caactctgtg	1680
	gggtggggag	gagcagatcc	aagttttgcg	gggtctaaag	ctgtgtgtgt	tgagggggat	1740
•	actctgttta	tgaaaaagaa	taaaaaacac	aaccacgaaa	aaaaaaaaa	aaaaaaaaa	1800
25	aaaaaaaaa	aaaaaaaccc	cccccgccc	ccccccctg	cag		1843
	<210> 50	•					
	<211> 11725	5					
	<212> DNA						
30	<213> Homo	sapiens					
	<400> 50						
•	agtgaatctg	ggcgagtaac	acaaaacttg	agtgtcctta	cctgaaaaat	agaggttaga	60
	gggatgctat	gtgccattgt	gtgtgtgtgt	tgggggtggg	gattgggggt	gatttgtgag	120
35	caattggagg	tgagggtgga	gcccagtgcc	cagcacctat	gcactgggga	cccaaaaagg	180
	agcatcttct	catgatttta	tgtatcagaa	attgggatgg	catgtcattg	ggacagcgtc	240
	ttttttcttg	tatggtggca	cataaataca	tgtgtcttat	aattaatggt	attttagatt	300
•	tgacgaaata	tggaatatta	cctgttgtgc	tgatcttggg	caaactataa	tatctctggg	360
	caaaaatgtc	cccatctgaa	aaacagggac	aacgttcctc	cctcagccag	ccactatggg	420
40·	gctaaaatga	gaccacatct	gtcaagggtt	ttgccctcac	ctecctccct	gctggatggc	480
	atccttggta	ggcagaggtg	ggcttcgggc	agaacaagcc	gtgctgagct	aggaccagga	540
	gtgctagtgc	cactgtttgt	ctatggagag	ggaggcctca	gtgctgaggg	ccaagcaaat	600
	atttgtggtt	atggattaac	tcgaactcca	ggctgtcatg	gcggcaggac	ggcgaacttg	660
	cagtatctcc	acgacccgcc	cctgtgagtc	ccctccagg	caggtctatg	aggggtgtgg	720
45						<b>.</b>	

agggaggct gccccggga gaagagact aggtggtgat gagggctgaa tcctccagcc 780

45

	agggtgctca	äcaagcctga	gcttggggta	aaaggacaca	aggccctcca	caggccaggc	840
	ctggcagcca	cagtctcagg	tccctttgcc	atgcgcctcc	ctctttccag	gccaagggtc	900
	cccaggccca	gggccattcc	aacagacagt	ttggagccca	ggaccctcca	ttctccccac	960
	cccacttcca	cctttggggg	tgtcggattt	gaacaaatct	cagaagcggc	ctcagaggga	1020
5 .	gtcggcaaga	atggagagca	gggtccggta	gggtgtgcag	aggccacgtg	gcctatccac	1080
	tggggagggt	tccttgatct	ctggccacca	gggctatctc	tgtggccttt	tggagcaacc	1140
	tggtggtttg	gggcaggggt	tgaatttcca	ggcctaaaac	cacacaggcc	tggccttgag	1200
	tcctggctct	gcgagtaatg	catggatgta	aacatggaga	cccaggacct	tgcctcagtc	1260
	ttccgagtct	ggtgcctgca	gtgtactgat	ggtgtgagac	cctactcctg	gaggatgggg	1320
10	gacagaatct	gatcgatccc	ctgggttggt	gacttccctg	tgcaatcaac	ggagaccagc	1380
	aagggttgga	tttttaataa	accacttaac	tcctccgagt	ctcagtttcc	ccctctatga	1440
	aatggggttg	acagcattaa	taactacctc	ttgggtggtt	gtgagcctta	actgaagtca	1500
	taatatctca	tgtttactga	gcatgagcta	tgtgcaaagc	ctgttttgag	agctttatgt	1560
	ggactaactc	ctttaattct	cacaacaccc	tttaaggcac	agatacacca	cgttattcca	1620
15	tccattttac	aaatgaggaa	actgaggcat	ggagcagtta	agcatcttgc	ccaacattgc	1680
	cctccagtaa	gtgctggagc	tggaatttgc	accgtgcagt	ctggcttcat	ggcctgccct	1740
	gtgaatcctg	taaaaattgt	ttgaaagaca	ccatgagtgt	ccaatcaacg	ttagctaata	1800
	ttctcagccc	agtcatcaga	ccggcagagg	cagccacccc	actgtcccca	gggaggacac	1860
	aaacatcctg	gcaccctctc	cactgcattc	tggagctgct	ttctaggcag	gcagtgtgag	1920
20	ctcagcccca	cgtagagcgg	gcagccgagg	ccttctgagg	ctatgtctct	agcgaacaag	1980
	gaccctcaat	tccagcttcc	gcctgacggc	cagcacacag	ggacagccct	ttcattccgc	2040
			agagcagcag				
	cctcctcaga	caggtgccag	tgcctccaga	atgtggcagc	tcacaagcct	cctgctgttc	2160
			cggcacacca				
25			ctacaaggcc				
			tgggtgaggt				
			agctgccaga				
			gcgatgacct				
20			tccctctcca				
30			actatccagc				
			acccctttcc				
			ggcacatggg		· -		
			gtttatctgc				
25			ccctcctcta				
35			ccctcctaac				
			gaccetteet				
			atggggaata				
			gcatgtggtg			-	
40			ccagggcttg				
40			atccccatgt	_			
			tcacacagag				
			tcctccaact				
			atgacctaga				
45			agetesests	•			
7.7	ttaggcccct		agctcccctc	CCCCCaaaac	cagactcagt	gttctccagc	3480







	ttttaatgtg	gaaattccta	tcttctgcct	ctagggcatt	tatcacttat	ttcttctaca	6240
	atctcccctt	tacttcctct	attttctctt	tctggacctc	ccattattca	gacctctttc	6300
	ctctagtttt	attgtctctt	ctatttccca	tctctttgac	tttgtgtttt	ctttcaggga	6360
	actttcttt	ttttctttt	ttttgagatg	gagtttcact	cttgttgtcc	caggctggag	6420
5	tgcaatgacg	tgatctcagc	tcaccacaac	ctccgcctcc	tggattcaag	cgattctcct	6480
	gccgcagcct	cccgagtagc	tgggattaca	ggcatgcgcc	accacgccca	gctaattttg	6540
	tgtttttagt	agagaagggg	tttctccgtg	ttggtcaagc	tggtcttgaa	ctcctgacct	6600
	caggtgatcc	acctgccttg	gcctcctaaa	gtgctgggat	tacaggcgtg	agccaccgcg	6660
	cccagcctct	ttcagggaac	tttctacaac	tttataattc	aattcttctg	cagaaaaaaa	6720
10	tttttggcca	ggctcagtag	ctcagaccaa	taattccagc	actttgagag	gctgaggtgg	6780
	gaggattgct	tgagcttggg	agtttgagac	tagcctgggc	aacacagtga	gaccctgtct	6840
	ctatttttaa	aaaaagtaaa	aaaagatcta	aaaatttaac	tttttattt	gaaataatta	6900
	gatatttcca	ggaagctgca	aagaaatgcc	tggtgggcct	gttggctgtg	ggtttcctgc	6960
	aaggccgtgg	gaaggccctg	tcattggcag	aaccccagat	cgtgagggct	ttccttttag	7020
15	gctgctttct	aagaggactc	ctccaagctc	ttggaggatg	gaagacgctc	acccatggtg	7080
	ttcggcccct	cagagcaggg	tggggcaggg	gagctggtgc	ctgtgcaggc	tgtggacatt	7140
	tgcatgactc	cctgtggtca	gctaagagca	ccactccttc	ctgaagcggg	gcctgaagtc	7200
	cctagtcaga	gcctctggtt	caccttctgc	aggcagggag	aggggagtca	agtcagtgag	7260
	gagggctttc	gcagtttctc	ttacaaactc	tcaacatgcc	ctcccacctg	cactgccttc	7320
20	ctggaagccc	cacagcctcc	tatggttdcg	tggtccagtc	cttcagcttc	tgggcgcccc	7380
	catcacgggc	tgagattttt	gctttccagt	ctgccaagtc	agttactgtg	tccatccatc	7440
	tgctgtcagc	ttctggaatt	gttgctgttg	tgccctttcc	attcttttgt	tatgatgcag	7500
	ctcccctgct	gacgacgtcc	cattgctctt	ttaagtctag	atatctggac	tgggcattca	7560
	aggcccattt	tgagcagagt	cgggctgacc	tttcagccct	cagttctcca	tggagtatgc	7620
25	gctctcttct	tggcagggag	gcctcacaaa	catgccatgc.	ctattgtagc	agctctccaa	7680
	gaatgctcac	ctccttctcc	ctgtaattcc	tttcctctgt	gaggagctca	gcagcatccc	7740
	attatgagac	cttactaatc	ccagggatca	ccccaacag	ccctgggġta	caatgagctt	7800
	ttaagaagtt	tạaccaccta	tgtaaggaga	cacaggcagt	gggcgatgct	gcctggcctg	7860
					tgactgactg		
30	gtaatttgta	tctcagggat	tacccccaac	agccctgggg	tacaatgagc	cttcaagaag	7980
					ctgcctggtc		
	cattcagtgg	çactgtttgt	tgactgactg	actgactgac	tggctgactg	gaggggttc	8100
					cttgaaccct		
2.5					cagtctcggg		
35		•		•	ccttgtggga		
					gaagaccaag		
					gacagcccct		
					tcactgagtc		
40					attgcgtttg		
40			•		ggggagagga		
					ggagggaggg		
					taggcagaag		
					ccagcctcca		
15					ataggaacca		
45	gtatttgccc	tggggactca	gactctgcaa	gggtcaggac	cccaaagacc	cggcagccca	8880

	gtgggaccac	agccaggacg	gcccttcaag	ataggggctg	agggaggcca	aggggaacat	8940
	ccaggcagcc	tgggggccac	aaagtcttcc	tggaagacac	aaggcctgcc	aagcctctaa	9000
	ggatgagagg	agctcgctgg	gcgatgttgg	tgtggctgag	ggtgactgaa	acagtatgaa	9060
	cagtgcagga	acagcatggg	caaaggcagg	aagacaccct	gggacaggct	gacactgtaa	9120
. 5	aatgggcaaa	aatagaaaac	gccagaaagg	cctaagccta	tgcccatatg	accagggaac	9180
	ccaggaaagt	gcatatgaaa	cccaggtgcc	ctggactgga	ggctgtcagg	aggcagccct	9240
	gtgatgtcat	catcccaccc	cattccaggt	ggtcctgctg	gactcaaaga	agaagctggc	9300
	ctgcggggca	gtgctcatcc	acccctcctg	ggtgctgaca	gcggcccact	gcatggatga	9360
	gtccaagaag	ctccttgtca	ggcttggtat	gggctggagc	caggcagaag	ggggctgcca	9420
10	gaggcctggg	tagggggacc	aggcaggctg	ttcaggtttg	ggggaccccg	ctccccaggt	9480
	gcttaagcaa	gaggcttctt	gagctccaca	gaaggtgttt	ggggggaaga	ggcctatgtg	9540
	ccccaccct	gcccacccat	gtacacccag	tattttgcag	tagggggttc	tctggtgccc	9600
	tcttcgaatc	tgggcacagg	tacctgcaca	cacatgtttg	tgaggggcta	cacagacctt	9660
	cacctctcca	ctcccactca	tgaggagcag	gctgtgtggg	cctcagcacc	cttgggtgca	9720
15	gagaccagca	aggcctggcc	tcagggctgt	gcctcccaca	gactgacagg	gatggagctg	9780
	tacagaggga	gccctagcat	ctgccaaagc	cacaagctgc	ttccctagca	ggctgggggc	9840
	tcctatgcat	tggccccgat	ctatggcaat	ttctggaggg	ggggtctggc	tcaactcttt	9900
	atgccaaaaa	gaaggcaaag	catattgaga	aaggccaaat	tcacatttcc	tacagcataa	9960
	tctatgccag	tggccccgtg	gggcttggct	tagaattccc	aggtgctctt	cccagggaac	10020
20	catcagtctg	gactgagagg	accttctctc	tcaggtggga	cccggccctg	tcctccctgg	10080
	cagtgccgtg	ttctgggggt	cctcctct	gggtctcact	gcccctgggg	tctctccagc	10140
	tacctttgct	ccatgttcct	ttgtggctct	ggtctgtgtc	tggggtttcc	aggggtctcg	10200
	ggcttccctg	ctgcccattc	cttctctggt	ctcacggctc	cgtgactcct	gaaaaccaac	10260
	cagcatccta	cccctttgga	ttgacacctg	ttggccactc	cttctggcag	gaaaagtcac	10320
25	cgttgatagg	gttccacggc	atagacaggt	ggctccgcgc	cagtgcctgg	gacgtgtggg	10380
	tgcacagtct	ccgggtgaac	cttcttcagg	ccctctccca	ggcctgcagg	ggcacagcag	10440
	tgggtgggcc	tcaggaaagt	gccactgggg	agaggctccc	cgcagcccac	tctgactgtg	10500
	ccctctgccc	tgcaggagag	tatgacctgc	ggcgctggga	gaagtgggag	ctggacctgg	10560
•	acatcaagga	ggtcttcgtc	caccccaact	acagcaagag	caccaccgac	aatgacatcg	10620
30		cctggcccag					
	cggacagcgg	ccttgcagag	cgcgagctca	atcaggccgg	ccaggagacc	ctcgtgacgg	10740
		ccacagcagc					
		gattcccgtg					
25	•	catgctgtgt		15			
35		gcccatggtc					
		gggctgtggg				_	
		gatccatggg					
		ctccctgcag					
40		agcacaccgg					
40		aacatttact					
		agcaactctg					
		gttgagggg					
		gccttttcca					
		accagettte					
45	cacagaggag	gaaactgagg	ggtctgaaag	gtttacatgg	tggagccagg	attcaaatct	11580

aggtctgact ccaaaaccca ggtgcttttt tctgttctcc actgtcctgg aggacagctg 11640 tttcgacggt gctcagtgtg gaggccacta ttagctctgt agggaagcag ccagagaccc 11700 agaaagtgtt ggttcagccc agaat 11725 5 <210> 51 <211> 32 <212> DNA <213> Artificial Sequence 10 <220> <223> Description of Artificial Sequence: Synthetic <400> 51 ttattttata tatataatat atatataaaa ta 32 15 <210> 52 <211> 9 <212> DNA <213> Artificial Sequence 20 <220> <223> Description of Artificial Sequence: Synthetic <400> 52 25 tataatata 9 <210> 53 <211> 16 <212> DNA 30 <213> Artificial Sequence <223> Description of Artificial Sequence: Synthetic 35 <400> 53 caatataaat atatag 16 <210> 54 <211> 44 40 <212> DNA <213> Artificial Sequence

<223> Description of Artificial Sequence: Synthetic

<220>

45

	W O 00/13219	PC1/0	1200/12/2
	<400> 54		
	tgtgtgtgta tgcgtgtgtg tagacacaca cgcatacaca cata		44
_	<210> 55		
5	<211> 41		
	<212> DNA		
	<213> Artificial Sequence		
	<220>		
10	<223> Description of Artificial Sequence: Synthetic		
	<400> 55		
	ttattttata tatataatat atataaaaa tatataata		41
15	<210> 56		
13			
	<211> 25 <212> DNA		
	<213> Artificial Sequence		
	(213) Arciffcial Sequence		
20	<220>		٠
	<223> Description of Artificial Sequence: Synthetic		
	<400> 55		
	tataatatac aatataaata tatag	٠	25
25			
	<210> 57		
	<211> 60		
•	<212> DNA		
	<213> Artificial Sequence		
30			
	<220>		
	<223> Description of Artificial Sequence: Synthetic		
	<400> 57		
35	caatataaat atatagtgtg tgtgtatgcg tgtgtgtaga cacacacgca	tacacacata	60
	<210> 58		
	•		
	<211> 101 <212> DNA		
40			
<del>-</del> -0	<213> Artificial Sequence		
	<220>		
	<223> Description of Artificial Sequence: Synthetic		

#### PCT/US00/15728

WO 00/75279 <400> 58 gtgtgtatgc gtgtgtgtag acacacacgc atacacacat a 5 <210> 59 <211> 723 <212> DNA <213> Homo sapiens 10 <400> 59 tttaggaaaa ttgttaagac tgggttgtat gcacactggt gtttattata ttatgtagtt 60 ttttctgtat ttttacaaca tttcagaatt aaaagcaaca gctagaaaaa gagggaaatg 120 gccgggtgca gtcgtcacgc ctgtaatccc agcactttgg gaggccaagg cgggcggatc 180 acgaggtcga gagatcgaga ccatcctggc caacatggtg aaaccccatc tctactaaaa 240 15 atacaaaaat taactgggca tggtggcatg cgcctgtagt cccaggagaa ttgcttgaac 300 ctgggaggeg gaggttgcag tgagccaaga teteaceact getetecage etggtgaeag 360 ggcaagactc cgtcaaaaaa aaaaaagaga gggagagcca gagtatgaaa aaggaagtca 420 gagecettta atgagteage tttgtaggte tecaggtagg aggetagtge tteagtgtet 480 aggacatagt aggtgttcag taaattaaat tcaggacaaa aagaacatgc cccaaggacc 540 20 atctgatate caettaaagt gatggactae etegttteee ttgtttatga atgggtteat 600 gcctaagact gtgtgcactt taatacaagg gcagtcgttc agaactagtc aggtcctgaa 660 aaggatttac caaatgttga gtgtgccctc tagtgttcac acttcccagc tttcttccta 720 taa 723 25 <210> 60 <211> 447 <212> DNA <213> Homo sapiens 30 <400> 60 aatatccagt tgagataatg gacttgcctc ttatctaata atacccaggc tcaatgcgtc 60 actgctttgt ccactttgcc caaaattcaa gcacagctaa gttgatattt taggacaaag 120 gcagcttact atccagccag aggggagtag aatatggtta agagagagtg gaaagaatga 180 atgageeetg etatteetea etgeetggat ggetataage acageeetta tggaggeett 240

35 aggtettget teacaatatt ceagtttgaa aagggtttga aaagacetee tagaaaaaate 300 agtagttttt ctcttttgag taacatgtag caaaaaaaat ttcatcatgt aggtacaggg 360 aacaccctag taactattaa tctcaaggag tcaagccagt gtgtttccta atgtatctgc 420 tgtatcccca tgaagcaaat tttgcca 447

40 <210> 61 <211> 103 <212> DNA <213> Homo sapiens

	<400> 61	
	aggtggatca aggcacttgc ttacaactgg aactgaaatc ctccaagtcg atctagacat	60
	tgagatggag aaaatattca ttgtcgactg taattatgca acg	103
5	<210> 62	
	<211> 7	
	<212> DNA	
	<213> Artificial Sequence	
10	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 62	
	gtggaag	7
15		
	<210> 63	
	<211> 7	•
	<212> DNA	
	<213> Artificial Sequence	
20		
	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 63	
25	gtggatg	7
	••	
	<210> 64 .	
	<211> 7	
	<212> DNA	
30	<213> Artificial Sequence	
	<220>	
	<223> Description of Artificial Sequence: Synthetic	
2.5		
35	<400> 64 .	
	ctggaag	7
	<210> 65	
40	<211> 7	
40	<212> DNA	
	<213> Artificial Sequence	
	.220.	
	<220>	
45	<223> Description of Artificial Sequence: Synthetic	
7.5		

	WO 00/75279		PCT/US00/15728
	<400> 65		•
	ctggatg		7
	<210> 66		
5	<211> 32		•
	<212> DNA		
	<213> Artificial Sequence		·
	<220>		
10	<223> Description of Artificial Sequence:	Synthetic	
	<400> 66 _		•
	acccattcag tcgaggaagg atagggtggt at		32
15	<210> 67	•	
	<211> 32		
	<212> DNA .		•
	<213> Artificial Sequence		
20	<220>		·
	<pre>&lt;223&gt; Description of Artificial Sequence:</pre>	Synthetic	•
	<400> 67		
	agccattgag tcgaggaagg atagggtggt at		32
25			
	<210> 68		
	<211> 32		
	<212> DNA	_	
30	<213> Artificial Sequence	•	•
	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 68		
35	agccattcag acgaggaagg atagggtggt at		32
	<210> 69		
	<211> 32		•
	<212> DNA		
40	<213> Artificial Sequence		
- =			
	<220>		
	<223> Description of Artificial Sequence:	Synthetic	

	WO 00/75279		PCT/US00/15728
	<400> 69		•
	agccattcag tçgaggaagg atagggtggt tt		. 32
	<210> 70		
5	<211> 32		
	<212> DNA		
	<213> Artificial Sequence		
	<220>		
10	<223> Description of Artificial Sequence:	Synthetic	
٠	<400> 70		
	agccattcag tcgaggatcc caagggtggt at		32
15	<210> 71		
	<211> 32		
	<212> DNA		
	<213> Artificial Sequence		
20	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 71		
	agccattcag tcgaggaagg atagggccta at		32 .
25			
	<210> 72		
	<211> 33		
	<212> DNA		
20	<213> Artificial Sequence	•	
30			
•	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
2.5	<400> 72		
35	agaccattca gtcgaggaag gatagggtgg tat		33
	<210> 73		
	<211> 33		
40	<212> DNA		
40	<213> Artificial Sequence		
	<220>		
	<223> Description of Artificial Sequence.	Synthetic	

	WO 00/75279	PCT/US00/15728
	<400> 73	
	agccattcag tcgaggaagg atagcggtgg tat	33
	<210> 74	
5	<211> 25	
-	<212> DNA	
	<213> Artificial Sequence	
10	<220>	
10	<223> Description of Artificial Sequence: Synthetic	
	· <400> 74	
	agccattcag tcgaggaagg ataat	25
15	<210> 75	
	<211> 28	
	<212> DNA	
	<213> Artificial Sequence	
20	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 75	
	·	28
25 ·	agccattcga ggaaggatag ggtggtat	26
	<210> 76	
	<211> 1273	
	<212> DNA	
	<213> Artificial Sequence	•
30	•	
	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 76	
35	ggcctctgac taactaatca ctttcccatc ttttgttaga tttgaatata taca	
	gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaat	•
	gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tcta	•
	cagecettga caaaattgtg aagttaaatt etecaetetg tecateagat acta	
40	tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tctt	
70	cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtac	
	tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca cagg tgagaggcta aaactcatca aaaacactac tccttttcct ctaccctatt cctc	
	ttacetttte caaateecaa teeccaaate agtttteet tttettacte eete	
	ttttacctc catggtcgtt aaaggagaga tggggagcat cattctgtta tact	
45	cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tqct	=

```
aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaaqtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
 5
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
10
      aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttaatccttc 1260
      tatcttgaat ctt
                                                                         1273
      <210> 77
      <211> 1273
15
      <212> DNA
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
20
      <400> 77
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
25
      cagcccttga caaaattgtg aagttaaatt ctccactctg tccatcagat actatggttc 240
      tocactatgg caactaactc actcaaattt coctocttag cagcattoca tottoccgat 300
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggcta aaactcatca aaaacactac tccttttcct ctaccctatt cctcaatctt 480
30
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
      ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600
      cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatata tatatatata aaatatataa tatacaatat aaatatatag 780
35
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
      agattggcat atcattgtaa ctaaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
40
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
      aagttgtoot titotggttt ogtgttoaco atggaacatt ttgattatag ttaatootto 1260
```

tatcttgaat ctt

```
<210> 78
      <211> 1273
      <212> DNA
       <213> Artificial Sequence
. 5
      <220>
      <223> Description of Artificial Sequence: Synthetic
       <400> 78
10
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
      cagccettga caaaattgtg aagttaaatt etceaetetg tecateagat actatggtte 240
      tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tcttcccgat 300
15
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggcta aaactcatca aaaacactac tccttttcct ctaccctatt cctcaatctt 480
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
      ttttaccctc catggtcgtt aaaggagaga tggggacgat cattctgtta tacttctgta 600
20
      cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
       agccatteta agagettgta tggttatgga ggtetgaeta ggeatgattt caegaaggea 900
-25
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
30
      aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttaatccttc 1260
      tatcttgaat ctt
                                                                         1273
      <210> 79
       <211> 1273
35
      <212> DNA
      <213> Artificial Sequence
       <223> Description of Artificial Sequence: Synthetic
40
       <400> 79
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatqt qttaggaaat tacagtcatt tctaagggcc 180
45
      cagecettga caaaattgtg aagttaaatt etecaetetg tecateagat actatggtte 240
```



```
tocactatgg caactaacto actoaatttt cootoottag cagoattoca tottocogat 300
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggeta aaacteatea aaaacactae teetttteet etaccetatt eeteaatett 480
 5
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
      ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600
      cacagitata catgictate aaacccagae tigeticeat agiggagaet tgetitteag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
10
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
      agattggcat atcattgtaa ctaaaaaaqc tqacattqac ccaqacatat tqtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
15
      aagaagttga agttgcctag accagaggac ataagtatgt actctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
      aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttaatccttc 1260
      tatcttgaat ctt
20
      <210> 80
      <211> 1275
      <212> DNA
      <213> Artificial Sequence
25
      <220>
      <223> Description of Artificial Sequence: Synthetic
      <400> 80
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
30
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
      cagecettga caaaattgtg aagttaaatt etecaetetg tecateagat actatggtte 240
      tecaetatgg caactaacte acteaatttt ceeteettag cageatteca tetteeegat 300
      ttcttctttg cttctccaac caaaacatca atgtttatta qttctgtata cagtacaqqa 360
35
      tetttggtet actetateae aaggeeagta ceacacteat gaagaaagaa cacaggagta 420
      gctgagaggc taaaactcat caaaaacact actccttttc ctctacccta ttcctcaatc 480
      ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc 540
      ccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg 600
      tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgctttc 660
40
      agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag 720
      ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat 780
      agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa 840
      taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg 900
      caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct 960
45
      ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac 1020
```



```
agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa 1080
      gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca 1140
      taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aaqtcattcc aatcagccaa 1200
      ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct 1260
 5
      tctatcttga atctt
                                                                         1275
      <210> 81
      <211> 1276
      <212> DNA
10
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
15
      <400> 81
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
      cagcccttga caaaattgtg aagttaaatt ctccactctg tccatcagat actatggttc 240
20
      tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tcttcccgat 300
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggcta aaactcatca aaaacactac tectttteet ctaccetatt cetcaatett 480
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
25
      ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600
      cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatataq 780
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
30
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
35
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
      gctaagttgt ccttttctgg tttcgtgttc accatggaac attttgatta tagttaatcc 1260
      ttctatcttg aatctt
                                                                         1276
      <210> 82
40
      <211> 1272
      <212> DNA
```

<213> Artificial Sequence

<220>

45

<223> Description of Artificial Sequence: Synthetic

```
<400> 82
5
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttitctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
      cagcccttga caaaattgtg aagttaaatt ctccactctg tccatcagat actatggttc 240
      tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tcttcccgat 300
10
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggeta aaacteatea aaaacactae teetttteet etaccetatt eeteaatett 480
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
      ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600
15
      cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
20
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
25
      aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttatccttct 1260
      atcttgaatc tt
                                                                         1272
      <210> 83
      <211> 1272
30
      <212> DNA
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
35
      <400> 83
      gcctctcact aactaatcac tttcccatct tttgttagat ttgaatatat acattctatg 60
      atcattgctt tttctcttta caggggagaa tttcatattt tacctgagca aattgattag 120
      aaaatggaac cactagagga atataatgtg ttaggaaatt acagtcattt ctaagggccc 180
40
      agcccttgac aaaattgtga agttaaattc tccactctgt ccatcagata ctatggttct 240
```

ccactatgge aactaactea etcaattte ecteettage ageatteeat etteegate 300 ttetttgett etceaaceaa aacateaatg tttattagtt etgtatacag tacaggatet 360 ttggtetaet etateacaag gecagtacea eacteatgaa gaaagaacae aggagtaget 420 gagaggetaa aacteateaa aaacactaet eettteete taceetatte etcaatettt 480

taccttttcc aaatcccaat ccccaaatca gtttttctct ttcttactcc ctctctccct 540



	tttaccctcc	atggtcgtta	aaggagagat	ggggagcatc	attctgttat	acttctgtac	600.
	acagttatac	atgtctatca	aacccagact	tgcttccata	gtggagactt	gcttttcaga	660
	acatagggat	gaagtaaggt	gcctgaaaag	tttgggggaa	aagtttcttt	cagagagtta	720
	agttatttta	tatatataat	atatatata	aatatataat	atacaatata	aatatatagt	780
5	gtgtgtgtgt	atgcgtgtgt	gtagacacac	acgcatacac	acatataatg	gaagcaataa	840
	gccattctaa	gagcttgtat	ggttatggag	gtctgactag	gcatgatttc	acgaaggcaa	900
	gattggcata	tcattgtaac	taaaaaagct	gacattgacc	cagacatatt	gtactctttc	960
			_			aatctacagc	
	- '					gagccaagca	
10					7	actagcatac	
				•		cagccaacta	
			gtgttcacca	tggaacattt	tgattatagt	taatccttct	
	atcttgaatc	tt					1272
15	<210> 84		•				
13	<211> 7				•		
	<212> DNA						
		ficial Seque	ence				
		•					
20	<220>			•	•		
	<223> Desci	ription of A	Artificial S	Sequence: S	Synthetic	•	
	<400> 84						•
	gcggaag						7
25							
•	<210> 85						
	<211> 1463						
	<212> DNA						
20	<213> Homo	sapiens					
30	-400- 05						•
	<400> 85	,	atatatacat		2200100000	tgacttgtta	60
	_	_				gcccagcatg	
					caccaccaca		
35		rcadcaddad	gcatctggtt	acaatcaaca	caagetgtte	cagccaattt	
					caagctgttc		
	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
	aaagaaactt ccaggatgtg	caggaggaat ccaccaatca	agggttttag taaggaggca	gagggcatgg ggggcctcct	ggaccctcct tccgctgctc	gcacccgaag cctgggactc	240 300
	aaagaaactt ccaggatgtg tctaggtgtc	caggaggaat ccaccaatca cgtggcctca	agggttttag taaggaggca gccccctct	gagggcatgg ggggcctcct gcacacctgc	ggaccetect teegetgete atetteette	gcacccgaag cctgggactc tcatcagctt	240 300 360
	aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt	caggaggaat ccaccaatca cgtggcctca aagcgtaaac	agggttttag taaggaggca gccccctct atggatgccc	gagggcatgg ggggcctcct gcacacctgc aggacctggc	ggaccetect teegetgete atetteette eteaatette	gcacccgaag cctgggactc	240 300 360 420
40	aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg	caggaggaat ccaccaatca cgtggcctca aagcgtaaac tactgacagt	agggttttag taaggaggca gccccctct atggatgccc gtgagaccct	gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga	ggaccetect teegetgete atetteette eteaatette teaateceet	gcacccgaag cctgggactc tcatcagctt cgagtctggt	240 300 360 420 480
40	aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg	caggaggaat ccaccaatca cgtggcctca aagcgtaaac tactgacagt caatcaatgg	agggttttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga	gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca	ggaccetect teegetgete atetteette eteaatette teaateceet catgeceegt	gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga	240 300 360 420 480 540
40	aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg agacttggag	caggaggaat ccaccaatca cgtggcctca aagcgtaaac tactgacagt caatcaatgg aaggaacgtg	agggttttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga ggcaagtctt	gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca cccaggaaca	ggaccetect teegetgete atetteette eteaatette teaateceet catgeceegt ggtagggcag	gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga ttagaggtgc	240 300 360 420 480 540
40	aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc	caggaggaat ccaccaatca cgtggcctca aagcgtaaac tactgacagt caatcaatgg aaggaacgtg tggtgcagcc	agggtttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga ggcaagtctt cggttcggag	gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca cccaggaaca caggaagacg	ggaccetect teegetgete atetteette eteaatette teaateceet catgeceegt ggtagggeag ettaataaat	gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga ttagaggtgc ggaggaaagg	240 300 360 420 480 540 600
	aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc tgcaggacac	caggaggaat ccaccaatca cgtggcctca aagcgtaaac tactgacagt caatcaatgg aaggaacgtg tggtgcagcc aggcaaaggt	agggtttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga ggcaagtctt cggttcggag gctgagctgg	gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca cccaggaaca caggaagacg accctttatt	ggaccetect teegetgete atetteette eteaatette teaateceet eatgeeeegt ggtagggeag ettaataaat tetgeeette	gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga ttagaggtgc ggaggaaagg gctgatagac	240 300 360 420 480 540 600 660 720
40 45	aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc tgcaggacac caccccggcc	caggaggaat ccaccaatca cgtggcctca aagcgtaaac tactgacagt caatcaatgg aaggaacgtg tggtgcagcc aggcaaaggt aggaaattgc	agggtttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga ggcaagtctt cggttcggag gctgagctgg tgcagccttt	gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca cccaggaaca caggaagacg accctttatt ctggaatccc	ggaccetect tecgetgete atctteette ctcaatette teaateceet catgeceegt ggtagggeag cttaataaat tetgecette gtteatttt	gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga ttagaggtgc ggaggaaagg gctgatagac tcccttctgg	240 300 360 420 480 540 600 660 720



	•		gtaacacaaa			•	
	_		attgtgtgtg				
	_		gtggagccca				
_			ttttatgtat				
5	gcgtcttttt	tcttgtatgg	tggcacataa	atacatgtgt	cttataatta	atggtatttt	1140
•	agatttgacg	aaatatggaa	tattacctgt	tgtgctgatc	ttgggcaaac	tataatatct	1200
	ctgggcaaaa	atgtccccat	ctgaaaaaca	gggacaacgt	tcctccctca	gccagccact	1260
	atggggctaa	aatgagacca	catctgtcaa	gggttttgcc	ctcacctccc	tccctgctgg	1320
	atggcatcct	tggtaggcag	aggtgggctt	cgggcagaac	aagccgtgct	gagctaggac	1380
10	caggagtgct	agtgccactg	tttgtctatg	gagaggagg	cctcagtgct	gagggccaag	1440
	caaatatttg	tggttatgga	tta			•	1463
	<210> 86					•	
	<211> 83						
15	<212> DNA						
	<213> Homo	sapiens					
		2					
	<400> 86						
		agtgccactg	tttgtctatg	gagagggagg	cctcagtgct	gagggccaag	60
20		tggttatgga		5-5-555-55		5-5555	83
	caaacacccg	0950000954					•
	<210> 87						
	<211> 614						
	<212> DNA						
25	<212> DNA <213> Homo	caniens					
23	CZIS> HOMO	sapiens	•				
	<400> 87						
		2002	25.55.55.55.55.55.55.55.55.55.55.55.55.5	aaaataaa	22655666	tanattatta	60
			atgtgtacct			-	
30	•		gttcctctag				
30			gcatctggtt				
			agggttttag				
			taaggaggca				
			gccccctct	_			
			atggatgccc				
35	•		gtgagaccct				
		_	aagccagcga				
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	ggggcatctc	tggt					614
40	<210> 88						
	<211> 661						
	<212> DNA				•		
,	<213> Homo	sapiens					

	<400> 88						
	gaattctgta	agcatttcct	atgtgtacct	gcccctgggc	aaggtgggcc	tgacttgtta	60
	gagtgttaga	gttttaccct	gttcctctag	gagggcctgg	taccaccaca	gcccagcatg	120
	gtgtggtgcc	tcagcaggag	gcatctggtt	acaatcaaca	caagctgttc	cagccaattt	180
5	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
	ccaggatgtg	ccaccaatca	taaggaggca	ggggcctcct	tccgctgctc	cctgggactc	300
	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
10	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	ggggcatctc	tggtgcagcc	cggttcggag	caggaagacg	cttaataaat	gctgatagac	660
	t				•		661
15	<210> 89						
	<211> 7	•					
	<212> DNA						
	<213> Homo	sapiens					
					·		
20	<400> 89						
	gaggaaa						7
					•		
	<210> 90	•					
25	<211> 7						
23	<212> DNA			•			
	<213> Homo	sapiens					
	<400> 90						
	caggaag					•	7
30	caggaag						′
50	<210> 91	•					
	<211> 102				•		
	<212> DNA						
	<213> Homo	sapiens					
35							
	<400> 91	•					
	tattttatat	atataatata	tatataaaat	atataatata	caatataaat	atatagtoto	60
		cgtgtgtgta					102
	J	3 3 3 3			<del>-</del>		
40	<210> 92						
	<211> 1380						
	<212> DNA						
	<213> Homo	sapiens					

	<400> 92						
	gaattctgta	agcatttcct	atgtgtacct	gcccctgggc	aaggtgggcc	tgacttgtta	60
	gagtgttaga	gttttaccct	gttcctctag	gagggcctgg	taccaccaca	gcccagcatg	120
	gtgtggtgcc	tcagcaggag	gcatctggtt	acaatcaaca	caagctgttc	cagccaattt	180
5	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
	ccaggatgtg	ccaccaatca	taaggaggca	ggggcctcct	tccgctgctc	cctgggactc	300
•	tctaggtgtc	cgtggcctca	gcccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
10	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	ggggcatctc	tggtgcagcc	cggttcggag	caggaagacg	cttaataaat	gctgatagac	660
	tgcaggacac	aggcaaaggt	gctgagctgg	accetttatt	tctgcccttc	tcccttctgg	720
	cacccggcc	aggaaattgc	tgcagccttt	ctggaatccc	gttcattttt	cttactggtc	780
15	cacaaaaggg	gccaaatgga	agcagcaaga	cctgagttca	aattaaatct	gccaactacc	840
	agctcagtga	atctgggcga	gtaacacaaa	acttgagtgt	ccttacctga	aaaatagagg	900
	ttagagggat	gctatgtgcc	attgtgtgtg	tgtgttgggg	gtggggattg	ggggtgattt	960
	gtgagcaatt	ggaggtgagg	gtggagccca	gtgcccagca	cctatgcact	ggggacccaa	1020
	aaaggagcat	cttctcatga	ttttatgtat	cagaaattgg	gatggcatgt	cattgggaca	1080
20	gcgtctttt	tcttgtatgg	tggcacataa	atacatgtgt	cttataatta	atggtatttt	1140
	agatttgacg	aaatatggaa	tattacctgt	tgtgctgatc	ttgggcaaac	tataatatct	1200
	ctgggcaaaa	atgtccccat	ctgaaaaaca	gggacaacgt	tcctccctca	gccagccact	1260
	atggggctaa	aatgagacca	catctgtcaa	gggttttgcc	ctcacctccc	tccctgctgg	1320
	atggcatcct	tggtaggcag	aggtgggctt	cgggcagaac	aagccgtgct	gagctaggac	1380
25							
	<210> 93						
	<211> 154						
	<212> DNA					•	
	<213> Homo	sapiens			•		
30							
	<400> 93						
				gttattttat			
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagcaataag	ccat	•		154
35		•					
	<210> 94	•					
	<211> 153					-	
	<212> DNA						
40	<213> Homo	sapiens					
40							
	<400> 94						
				ttattttata			
				gtgtgtgtat	gcgtgtgtgt	agacacacac	
45	gcatacacac	atataatgga	agcaataagc	cat			153
45							

	<210> 95						
•	<211> 152						
	<212> DNA						
	<213> Homo	sapiens					
5							
	<400> 95						
	gggggaaaag	tttctttcag	agagttaagt	tattttatat	atataatata	tatataaaat	60
	atataatata	castataeat	atatagtgtg	tgtgtgtatg	cgtgtgtgta	gacacacacg	120
	catacacaca	tataatggaa	gcaataagcc	at			152
10							
	<210> 96						
	<211> 151	•					
	<212> DNA						
	<213> Homo	sapiens					-
15							
	<400> 96						
	ggggaaaagt	ttctttcaga	gagttaagtt	attttatata	tataatatat	atataaaata	60
	tataatatac	aatataaata	tatagtgtgt	gtgtgtatgc	gtgtgtgtag	acacacacgc	120
	atacacacat	ataatggaag	caataagcca	t			151
20							
	<210> 97						
	<211> 150						
	<212> DNA						•
	<213> Homo	sapiens					
25							
	<400> 97						
	gggaaaagtt	tctttcagag	agttaagtta	ttttatatat	ataatatata	tataaaatat	60
٠.	ataatataca	atataaatat	atagtgtgtg	tgtgtatgcg	tgtgtgtaga	cacacacgca	120
	tacacacata	taatggaagc	aataagcca't				150
30						-	
	<210> 98						
	<211> 149						
	<212> DNA						
	<213> Homo	sapiens					
35							
	<400> 98					•	
	ggaaaagttt	ctttcagaga	gttaagttat	tttatatata	taatatatat	ataaaatata	60
	taatatacaa	tataaatata	tagtgtgtgt	gtgtatgcgt	gtgtgtagac	acacacgcat	120
	acacacatat	aatggaagca	ataagccat				149
40						·	•
	<210> 99	•	•				
	<211> 148						
	<212> DNA						
	<213> Homo	sapiens					
45			•			•	•

	<400> 99						
	gaaaagtttc	tttcagagag	ttaagttatt	ttatatatat	aatatatata	taaaatatat	60 ·
	aatatacaat	ataaatatat	agtgtgtgtg	tgtatgcgtg	tgtgtagaca	cacacgcata	120
	cacacatata	atggaagcaa	taagccat				148
5							
	<210> 100						
	<211> 147						
	<212> DNA			•			
	<213> Homo	sapiens			•		
10		_					
	<400> 100						
	aaaagtttct	ttcagagagt	taagttattt	tatatatata	atatatatat	aaaatatata	60
		taaatatata					
		tggaagcaat					147
15		55 5	_				
	<210> 101						
	<211> 146	•					
	<212> DNA		•				
	<213> Homo	sapiens					
20		•					
	<400> 101						
	aaagtttctt	tcagagagtt	aagttatttt	atatatata	tatatatata	aaatatataa	60
	_	zaatatatag	_				
		ggaagcaata			-	_	146
25					•		
	<210> 102						
	<211> 145		•				
	<212> DNA						
	<213> Homo	sapiens					
30				•			
	<400> 102						
	aagtttcttt	cagagagtta	agttattta	tatatataat	atatatata	aatatataat	60
	atacaatata	aatatatagt	gtgtgtgtgt	atgcgtgtgt	gtagadacad	acgcatacac	120
	acatataatg	gaagcaataa	gccat	•			145
35						•	
	<210> 103				•		
	<211> 144						
	<212> DNA	•					
	<213> Homo	sapiens					
40							
	<400> 103						
	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	atatataata	60
	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	cgcatacaca	120
	catataatgg	aagcaataag	ccat		•		144
45							

	<210>	104						
	<211>	143	•					
	<212>	DNA						
	<213>	Homo	sapiens					
5								
	<400>	104						
	gtttct	ttca	gagagttaag	ttattttata	tatataatat	atatataaaa	tatataatat	60
	acaata	ataaa	tatatagtgt	gtgtgtgtat	gcgtgtgtgt	agacacacac	gcatacacac	120
	atataa	atgga	agcaataagc	cat				143
10				·		•		
	<210>	105						
	<211>	142						
	<212>	DNA						
	<213>	Homo	sapiens					
15								
	<400>	105						
	tttctt	tcag	agagttaagt	tattttatat	atataatata	tatataaaat	atataatata	60
	caatat	aaat	atatagtgtg	tgtgtgtatg	cgtgtgtgta	gacacacacg	catacacaca	120
	tataat	ggaa	gcaataagcc	at				142
20	•							
	<210>	106						
	<211>	141						
	<212>	DNA	•					•
	<213>	Homo	sapiens					
25								
	<400>	106				•		
	ttcttt	caga	gagttaagtt	attttatata	tataatatat	atataaaata	tataatatac	60
•	aatata	aaata	tatagtgtgt	gtgtgtatgc	gtgtgtgtag	acacacacgc	atacacacat	120
	ataato	ggaag	caataagcca	t				141
30						•		
	<210>	107						
	<211>	140				•		
	<212>	DNA		•	•			
	<213>	Homo	sapiens					
35								
	<400>							
			_				ataatataca	
				tgtgtatgcg	tgtgtgtaga	cacacacgca	tacacacata	120
	taatg	gaagc	aataagccat		•			140
40								
	<210>							
	<211>							
	<212>							
4.5	<213>	Homo	sapiens			•		
45			•	•				

- 60 -

	<400> 108	•					
	ctttcagaga	gttaagttat	tttatatata	taatatatat	ataaaatata	taatatacaa	60
	tataaatata	tagtgtgtgt	gtgtatgcgt	gtgtgtagac	acacacgcat	acacacatat	120
	aatggaagca						139
5							
	<210> 109						
	<211> 138						
	<212> DNA						
	<213> Homo	sapiens					
10	•	_					
	<400> 109						
	tttcagagag	ttaagttatt	ttatatatat	aatatatata	taaaatatat	aatatacaat	60
	_				cacacgcata		
	atggaagcaa		3 3, 3	3 3 3			138
15	33 3	J					
	<210> 110						
	<211> 137						
	<212> DNA		•				
	<213> Homo	sapiens					
20		•					
	<400> 110	-					
	ttcagagagt	taagttattt	tatatatata	atatatatat	aaaatatata	atatacaata	60
	taaatatata	gtgtgtgtgt	gtatgcgtgt	gtgtagacac	acacgcatac	acacatataa	120
	tggaagcaat				_		137
25.					•		
	<210> 111						
	<211> 136					•	
	<212> DNA						
	<213> Homo	sapiens					
30							
	<400> 111						•
	tcagagagtt	aagttatttt	atatatata	tatatata	aaatatataa	tatacaatat	60
	aaatatatag	tgtgtgtgtg	tatgcgtgtg	tgtagacaca	cacgcataca	cacatataat	120
	ggaagcaata	agccat					136
35							
	<210> 112						
	<211> 135						
	<212> DNA						
	<213> Homo	sapiens					
40							
	<400> 112						
	cagagagtta	agttatttta	tatatataat	atatatataa	aatatataat	atacaatata	60
	aatatatagt	gtgtgtgtgt	atgcgtgtgt	gtagacacac	acgcatacac	acatataatg	120
	gaagcaataa	gccat					135
45							

WO 00/75279 PCT/US00/15728 <210> 113 <211> 134 <212> DNA <213> Homo sapiens 5 <400> 113 atatatagtg tgtgtgta tgcgtgtgtg tagacacaca cgcatacaca catataatgg 120 aagcaataag ccat 134 10 <210> 114

<211> 133 <212> DNA <213> Homo sapiens

<400> 114

15

20

25

. 35

40

45

tatatagtgt gtgtgtgtat gcgtgtgtgt agacacacac gcatacacac atataatgga 120 agcaataagc cat 133

<210> 115 <211> 132

<212> DNA

<213> Homo sapiens

<400> 115

atatagtgtg tgtgtgtatg cgtgtgtgta gacacacag catacacaca tataatggaa 120 gcaataagcc at 132

30 <210> 116 <211> 131 <212> DNA

<213> Homo sapiens

<400> 116

tatagtgtgt gtgtgtatgc gtgtgtgtag acacacacgc atacacacat ataatggaag 120 caataagcca t 131

<210> 117 <211> 130 <212> DNA <213> Homo sapiens

	<400> 117						
	agttaagtta	ttttatatat	ataatatata	tataaaatat	ataatataca	atataaatat	60
						taatggaagc	
	aataagccat						130
5							
	<210> 118						
	<211> 129						
	<212> DNA				•		
	<213> Homo	sapiens					
10							
	<400> 118						
	gttaagttat	tttatatata	taatatatat	ataaaatata	taatatacaa	tataaatata	60
						aatggaagca	
	ataagccat		<del>-</del>				129
15							
	<210> 119						
	<211> 128						
	<212> DNA						
	<213> Homo	sapiens					
20							
	<400> 119						
	ttaagttatt	ttatatatat	aatatatata	taaaatatat	aatatacaat	ataaatatat	60
	agtgtgtgtg	tgtatgcgtg	tgtgtagaca	cacacgcata	cacacatata	atggaagcaa	120
	taagccat						128
25					•		
	<210> 120						
	<211> 127						
	<212> DNA					•	
	<213> Homo	sapiens				•	
30		•					
	<400> 120		•				
	taagttattt	tatatatata	atatatatat	aaaatatata	atatacaata	taaatatata	60
	gtgtgtgtgt	gtatgcgtgt	gtgtagacac	acacgcatac	acacatataa	tggaagcaat	120
	aagccat			•			127
35						•	
	<210> 121						
	<211> 126						
	<212> DNA						
	<213> Homo	sapiens					•
40		•					
	<400> 121						
	aagttatttt	atatataa	tatatata	aaatatataa	tatacaatat	aaatatatag	60
	tgtgtgtgtg	tatgcgtgtg	tgtagacaca	cacgcataca	cacatataat	ggaagcaata	120
	agccat						126
45		•					

WO 00/75279

<210> 122
<211> 125
<212> DNA
<213> Homo sapiens

5 <400> 122

agttatttta tatatataat atatataaa aatatataat atacaatata aatatatagt 60 gtgtgtgtgt atgcgtgtgt gtagacacac acgcatacac acatataatg gaagcaataa 120 gccat

10 <210> 123

<210> 123 <211> 124 <212> DNA <213> Homo sapiens

<400> 123

20

15

<210> 124 <211> 123 · <212> DNA <213> Homo sapiens

25

30

<210> 125 <211> 153 <212> DNA <213> Homo sapiens

35

<400> 125

ttgggggaaa agtttctttc agagagttaa gttattttat atatataata tatataaaa 60
atatataata tacaatataa atatatagtg tgtgtgtgta tgcgtgtgtg tagacacaca 120
cgcatacaca catataatgg aagcaataag cca 153

40

<210> 126 <211> 152 <212> DNA <213> Homo sapiens

	<400> 1	26	•				•	
	ttggggg	aaa	agtttettte	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatata	ata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatac	aca	catataatgg	aagcaataag	cc			152
5								
	<210> 1	27						
	<211> 1	51						
	<212> D	NA						
	<213> H	omo	sapiens					
10							•	
	<400> 1	27						
	ttggggg	aaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatata	ata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgċatac	aca	catataatgg	aagcaataag	С			151
15								
	<210> 1	28						
	<211> 1	50						•
	<212> D	NA						
	<213> H	omo	sapiens					
20				•				
	<400> 1	28						
	ttggggg	aaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatata	ata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatac	aca	catataatgg	aagcaataag				150
25								
	<210> 1	29						
	<211> 1	49						
	<212> D	NA					-	
	<213> He	omo	sapiens					
30								
	<400> 1	29				_		
	ttggggg	aaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatata	ata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatac	aca	catataatgg	aagcaataa				149
35								
•	<210> 1	30						
	<211> 1	48		•				
	<212> Di	NA		•				
	<213> He	omo	sapiens					
40							•	
	<400> 13	30						
			_		_		tatatataaa	
			_		tgtgtgtgta	tgcgtgtgtg	tagacacaca	12.0
4.5	cgcatac	aca	catataatgg	aagcaata				148
45			•					

	<210> 131					•	
	<211> 147						
	<212> DNA						
	<213> Homo	sapiens.					
5							
	<400> 131	•					
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagcaat				147
10							
	<210> 132						
	<211> 146						
	<212> DNA						
	<213> Homo	sapiens	٠		•		
13							
	<400> 132						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagcaa				146
20				•			
	<210> 133						
	<211> 145						
	<212> DNA						
	<213> Homo	sapiens					
25							
	<400> 133						
	ttgggggaaa	agtttcttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagca				145
30							
	<210> 134						
	<211> 144						
•	<212> DNA		•				
	<213> Homo	sapiens					
35		•			-		
	<400> 134						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagc				144
40							
	<210> 135						
	<211> 143	•					
	<212> DNA						
	<213> Homo	sapiens					
45							

	<400> 135						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aag				143
5		•					
	<210> 136	•					
	<211> 142			•			
	<212> DNA		•				
	<213> Homo	sapiens					
10							
	<400> 136						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aa				142
15							
	<210> 137						
	<211> 141						
	<212> DNA						
	<213> Homo	sapiens					
20							
	<400> 137						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	a				141
25							
	<210> 138			•			
	<211> 140						
	<212> DNA						
	<213> Homo	sapiens					
30							•
	<400> 138	•					
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg					140
35							
	<210> 139						
	<211> 139						
	<212> DNA						
	<213> Homo	sapiens					
40							
	<400> 139						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatg					139
45							

	<210> 140						
	<211> 138					•	
	<212> DNA						
	<213> Homo	sapiens					
5							
	<400> 140						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataat					138
10							
	<210> 141						
	<211> 137					•	
	<212> DNA						
	<213> Homo	sapiens					
15							
	<400> 141						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataa					137
20							
	<210> 142						
	<211> 136			•			
	<212> DNA						
	<213> Homo	sapiens	•				
25 .							
	<400> 142						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catata					136
30		•					
	<210> 143					•	
	<211> 135	•					
	<212> DNA						
	<213> Homo	sapiens			•	•	
35							
	<400> 143						
				gttattttat			
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
40	cgcatacaca	catat					135
40							
	<210> 144						
	<211> 134	•					
	<212> DNA			•	•		
15	<213> Homo	sapiens			·		
45							

```
<400> 144
      atatataata tacaatataa atatatagtg tgtgtgtgta tgcgtgtgtg tagacacaca 120
      cgcatacaca cata
                                                                       134
 5
      <210> 145
      <211> 659
      <212> DNA
      <213> Homo sapiens
10
      <400> 145
      gaattetgta agcattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
15
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggigte egiggeetea geceeetet geacacetge atetteette teateageti 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccet actectetga teaateceet gggttggtga 480
20
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gctgataga 659
      <210> 146
25
      <211> 658
      <212> DNA
      <213> Homo sapiens
      <400> 146
30
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatq 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
35
      totaggtgtc cgtggcctca gcccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
40
      ggggcatete tggtgcagee eggtteggag caggaagaeg ettaataaat getgatag
      <210> 147
      <211> 657
      <212> DNA
45
      <213> Homo sapiens
```

```
<400> 147
      gaattetgta ageattteet atgtgtaeet geecetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      ettecetgtg caatcaatgg aagccagega ggcagggtea catgeceegt ttagaggtge 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gctgata
      <210> 148
15
      <211> 656
      <212> DNA
      <213> Homo sapiens
      <400> 148
20
      gaattetgta ageattteet atgtgtacet geeectggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
25
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtqc 540°
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
30
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gctgat
                                                                         656
      <210> 149
      <211> 655
      <212> DNA
35
      <213> Homo sapiens
      <400> 149
      gaattotgta agcatttoot atgtgtacot goccotgggo aaggtgggoo tgacttgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
40
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
45
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
```

```
cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee cggtteggag caggaagaeg ettaataaat getga
 5
      <210> 150
      <211> 654.
      <212> DNA
      <213> Homo sapiens
10
      <400> 150
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
15
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcaqctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
20
      agacttggag aaggaacgtg ggcaagtett cecaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee eggtteggag caqqaaqaeq ettaataaat qetq
      <210> 151
      <211> 653
25
      <212> DNA
      <213> Homo sapiens
      <400> 151
      gaattetgta agcattteet atgtgtaeet geeettggge aaggtgggee tgaettgtta 60
30
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
35
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gct
40
      <210> 152*
      <211> 652
      <212> DNA
      <213> Homo sapiens
```

```
<400> 152
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaaq 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgqgactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cetetgettt aagegtaaae atggatgeee aggaeeétgge etcaatette egagtetggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtett eecaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee eggtteggag caggaagaeg ettaataaat qe
      <210> 153
15
      <211> 651
      <212> DNA
      <213> Homo sapiens
      <400> 153
20
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatq 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
25
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccet actectetga teaateceet gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
30
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat g
                                                                         651
      <210> 154
      <211> 650
      <212> DNA
35
      <213> Homo sapiens
      <400> 154
      gaattetgta ageattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca qcccaqcatq 120
40
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgto cgtggcotca gccccctct gcacacctgc atottccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
45
      acttatggtg tactgacagt gtgagaccet acteetetga teaateeeet gggttggtga 480
```

```
cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee eggtteggag caggaagaeg ettaataaat
5
      <210> 155
      <211> 649
      <212> DNA
      <213> Homo sapiens
10
      <400> 155
      gaattetgta agcattteet atgtgtacet geeectggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
15
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
20
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee cggttcggag caggaagaeg ettaataaa
      <210> 156
      <211> 648
25
      <212> DNA
      <213> Homo sapiens
      <400> 156
      gaattetgta ageattteet atgtgtaeet geeectggge aaggtgggee tgaettgtta 60
30
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
35
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataa
40
      <210> 157
      <211> 647
      <212> DNA
      <213> Homo sapiens
45
```

```
<400> 157
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca qcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaaq 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgto cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagegtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaata
                                                                          647
      <210> 158
15
      <211> 646
      <212> DNA
      <213> Homo sapiens
      <400> 158
20
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
25
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480 .
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
30
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaat
                                                                         646
      <210> 159
      <211> 645
      <212> DNA
35
      <213> Homo sapiens
      <400> 159
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
40
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaaq 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cqaqtctqqt 420
45
      acttatggtg tactgacagt gtgagaccet acteetetga teaateeeet gggttggtga 480°
```

```
cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
       ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaa
 5
       <210> 160
       <211> 644
       <212> DNA
       <213> Homo sapiens
10
       <400> 160
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
15
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cetetgettt aagegtaaac atggatgeec aggacetgge etcaatette egagtetggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
20
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg ctta
                                                                         644
      <210> 161
      <211> 643
25
      <212> DNA
      <213> Homo sapiens
      <400> 161
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
30
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggceteet teegetgete eetgggaete 300
      totaggtgto ogtggootca gooccootct goacacotgo atottootto toatcagett 360
35
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cqaqtctqqt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg ctt
                                                                         643
40
      <210> 162
      <211> 642
      <212> DNA
      <213> Homo sapiens
45
```

	<400> 162						
		agcatttcct	atototacct	accetagae	aaggtgggc	tgacttgtta	60
						gcccagcatg	
						cagccaattt	
5						gcacccgaag	
-						cctgggactc	
						tcatcagctt	
						cgagtctggt	
						gggttggtga	
10				•		ttagaggtgc	
					_	ggaggaaagg	
				caggaagacg		55-5555	642
	5555	0330300300	-5555-5				<b>V</b> 12
	<210> 163		•				
15	<211> 641						
	<212> DNA						
	<213> Homo	sapiens					•
	<400> 163						
20	gaattctgta	agcatttcct	atgtgtacct	gcccctgggc	aaggtgggcc	tgacttgtta	60
	gagtgttaga	gttttaccct	gttcctctag	gagggcctgg	taccaccaca	gcccagcatg	120
	gtgtggtgcc	tcagcaggag	gcatctggtt	acaatcaaca	caagctgttc	cagccaattt	180
	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
	ccaggatgtg	ccaccaatca	taaggaggca	ggggcctcct	tccgctgctc	cctgggactc	300
25	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
30	ggggcatctc	tggtgcagcc	cggttcggag	caggaagacg	C		641
	<210> 164						
	<211> 640			•			
	<212> DNA						
35	<213> Homo	sapiens					
	<400> 164						
						tgacttgtta	
40						gcccagcatg	
40						cagccaattt	
						gcacccgaag	
						cctgggactc	
						tcatcagctt	
45						cgagtctggt	
70	acttatggtg	Lactgacagt	gtgagaccct	actcctctga	ccaatcccct	gggttggtga	480

```
cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
       ggggcatctc tggtgcagcc cggttcggag caggaagacg
 5
       <210> 165
       <211> 639
       <212> DNA
       <213> Homo sapiens
10
       <400> 165
       gaattctgta agcatttcct atgtgtacct gcccctgggc aaggtgggcc tgacttgtta 60
       gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
       gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
       aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
15
       ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
       cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
       acttatggtg tactgacayt gtgagaccct actcctctga tcaatcccct gggttggtga 480
       cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
20
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
       ggggcatete tggtgcagee eggtteggag caggaagae
       <210> 166
       <211> 638
25
       <212> DNA
       <213> Homo sapiens
       <400> 166
       gaattetgta agcattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
30
       gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
       gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
       aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
       ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
35
       cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
       acttatggtg tactgacayt gtyagaccct actcctctga tcaatcccct gggttggtga 480
       cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
       ggggcatctc tggtgcagcc cggttcggag caggaaga
                                                                          638
40
       <210> 167
       <211> 637
       <212> DNA
       <213> Homo sapiens
```

```
<400> 167
      gaattetgta ageattteet atgtgtacet geceetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
       aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
       ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
       cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
       acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee eggtteggag eaggaag
                                                                         637
       <210> 168
15
       <211> 650
       <212> DNA
       <213> Homo sapiens
      <400> 168
20
       agcatttcct atgtgtacct gcccctgggc aaggtgggcc tgacttgtta gagtgttaga 60
      gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg gtgtggtgcc 120
      tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt aaagaaactt 180
       caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag ccaggatgtg 240
       ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc tctaggtgtc 300
25
       egtggeetea geceeetet geacacetge atetteette teateagett eetetgettt 360
       aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt acttatggtg 420
       tactgacagt gtgagaccct actectetga teaateceet gggttggtga ettecetgtg 480
       caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc agacttggag 540
       aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg ggggcatctc 600
30
       tggtgcagcc cggttcggag caggaagacg cttaataaat gctgatagac
                                                                         650
       <210> 169
       <211> 640
       <212> DNA
35
       <213> Homo sapiens
       <400> 169
      atgtgtacct gcccctgggc aaggtgggcc tgacttgtta gagtgttaga gttttaccct 60
      gttcctctag gagggcctgg taccaccaca gcccagcatg gtgtggtgcc tcagcaggag 120
40
      gcatctggtt acaatcaaca caagctgttc cagccaattt aaagaaactt caggaggaat 180
       agggttttag gagggcatgg ggaccctcct gcacccgaag ccaggatgtg ccaccaatca 240
       taaggaggca ggggcctcct tccgctgctc cctgggactc tctaggtgtc cgtggcctca 300
      gececetet geacacetge atetteette teateagett cetetgettt aagegtaaac 360
       atggatgccc aggacctggc ctcaatcttc cgagtctggt acttatggtg tactgacagt 420
45
      gtgagaccct actcctctga tcaatcccct gggttggtga cttccctgtg caatcaatgg 480
```



	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	agacttggag	aaggaacgtg	540
	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	ggggcatctc	tggtgcagcc	600
	cggttcggag	caggaagacg	cttaataaat	gctgatagac			640
					•		
5	<210> 170						
	<211> 610						
	<212> DNA					•	
	<213> Homo	sapiens					
10 -	<400> 170						
	tgacttgtta	gagtgttaga	gttttaccct	gttcctctag	gagggcctgg	taccaccaca	60
				gcatctggtt			
	cagccaattt	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	180
				taaggaggca			
15	cctgggactc	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	300
	tcatcagctt	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	360
	cgagtctggt	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	420
	gggttggtga	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	480
	ttagaggtgc	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	540
20	ggaggaaagg	ggggcatctc	tggtgcagcc	cggttcggag	caggaagacg	cttaataaat	600
	gctgatagac						610
	<210> 171						
0.5	<211> 300						
25	<212> DNA						
	<213> Homo	sapiens				•	
	.400. 177						
	<400> 171						<b>50</b>
30		aagcgtaaac	arggargeee	aggaccugge	CLCaalCitc	cgagtctggt	60
20							120
			gtgagaccct	actectetga	tcaatcccct	gggttggtga	
	cttccctgtg	caatcaatgg	gtgagaccct aagccagcga	actectetga ggeagggtea	tcaatcccct catgccccgt	gggttggtga ttagaggtgc	180
	cttccctgtg agacttggag	caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtca cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	180 240
	cttccctgtg agacttggag	caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggeagggtea	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	180 240
35	cttccctgtg agacttggag ggggcatctc	caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtca cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	180 240
35	cttccctgtg agacttggag ggggcatctc <210> 172	caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtca cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	180 240
35	cttccctgtg agacttggag ggggcatctc <210> 172 <211> 100	caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtca cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	180 240
35	cttccctgtg agacttggag ggggcatctc <210> 172 <211> 100 <212> DNA	caatcaatgg aaggaacgtg tggtgcagcc	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtca cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	180 240
35	cttccctgtg agacttggag ggggcatctc <210> 172 <211> 100	caatcaatgg aaggaacgtg tggtgcagcc	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtca cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	180 240
	cttccctgtg agacttggag ggggcatctc <210> 172 <211> 100 <212> DNA <213> Homo	caatcaatgg aaggaacgtg tggtgcagcc	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtca cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	180 240
35	cttccctgtg agacttggag ggggcatctc <210> 172 <211> 100 <212> DNA <213> Homo <400> 172	caatcaatgg aaggaacgtg tggtgcagcc . sapiens	gtgagaccct aagccagcga ggcaagtctt cggttcggag	actectetga ggcagggtca cccaggaaca caggaagacg	tcaatcccct catgccccgt ggtagggcag cttaataaat	gggttggtga ttagaggtgc ggaggaaagg gctgatagac	180 240 300
	cttccctgtg agacttggag ggggcatctc <210> 172 <211> 100 <212> DNA <213> Homo <400> 172 ggcaagtctt	caatcaatgg aaggaacgtg tggtgcagcc . sapiens	gtgagaccct aagccagcga ggcaagtctt cggttcggag	actcctctga ggcagggtca cccaggaaca caggaagacg	tcaatcccct catgccccgt ggtagggcag cttaataaat	gggttggtga ttagaggtgc ggaggaaagg gctgatagac	180 240 300

```
<210> 173
      <211> 71.
      <212> DNA
      <213> Homo sapiens
 5
      <400> 173
      gggaggaaag gggggcatct ctggtgcagc ccggttcgga gcaggaagac gcttaataaa 60
      tgctgataga c
10
      <210> 174
      <211> 613
      <212> DNA
      <213> Homo sapiens
15
      <400> 174
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
20
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geececetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
25
      agacttggag aaggaacgtg ggcaagtett cecaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tgg
                                                                         613
      <210> 175
      <211> 612
30
      <212> DNA
      <213> Homo sapiens
      <400> 175
      gaattetgta agcattteet atgtgtacet geeectggge aaggtgggee tgaettgtta 60
35
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
40
      cctctgcttt aagegtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tg
                                                                         612
45
```

```
<210> 176
       <211> 611
       <212> DNA
       <213> Homo sapiens
 5
       <400> 176
       gaattetgta ageattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
       gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
10
       aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
15
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc t
                                                                          611
      <210> 177
20
      <211> 610
       <212> DNA
      <213> Homo sapiens
      <400> 177
25
      gaattotgta agoatttoot atgtgtacot goocotgggo aaggtgggoo tgaottgttá 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccaqcatq 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
30
      totaggtgto ogtggcotca goodcootct goadacotgo atottootto toatcagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtett eecaggaaca ggtagggcag ggaggaaagg 600
35
      ggggcatctc
                                                                         610
      <210> 178
      <211> 609
      <212> DNA
40
      <213> Homo sapiens
      <400> 178
      gaattetgta ageattteet atgtgtaeet geeettggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
45
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
```

```
aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cqtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
5
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatct
10
      <210> 179
      <211> 608
      <212> DNA
      <213> Homo sapiens
15
      <400> 179
      gaattetgta ageattteet atgtgtaeet geeectggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
20
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
25
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
                                                                         608
      ggggcatc
      <210> 180
      <211'> 607
30
      <212> DNA
      <213> Homo sapiens
      <400> 180
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
35
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
40
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
                                                                         607
      ggggcat
```

```
<210> 181
       <211> 606
       <212> DNA
      <213> Homo sapiens
 5
       <400> 181
      gaattetgta ageattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
10
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccet acteetetga teaateeeet gggttggtga 480
15
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggca
                                                                         606
      <210> 182
20
      <211> 605
      <212> DNA
      <213> Homo sapiens
      <400> 182
25
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
30
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccet acteetetga teaateeeet gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
35
                                                                         605
      ggggc
      <210> 183
      <211> 604
      <212> DNA
40
      <213> Homo sapiens
      <400> 183
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
45
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
```

						•	
		caggaggaat					
		ccaccaatca					
	tctaggtgtc	cgtggcctca	gcccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
_	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
5	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	<b>a</b> aaa						604
						•	
10	<210> 184						
	<211> 603						
	<212> DNA						
	<213> Homo	sapiens					
15	<400> 184						
	gaattctgta	agcatttcct	atgtgtacct	gcccctgggc	aaggtgggcc	tgacttgtta	60
	gagtgttaga	gttttaccct	gttcctctag	gagggcctgg	taccaccaca	gcccagcatg	120
	gtgtggtgcc	tcagcaggag	gcatctggtt	acaatcaaca	caagctgttc	cagccaattt	180
	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
20	ccaggatgtg	ccaccaatca	taaggaggca	ggggcctcct	tccgctgctc	cctgggactc	300
	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
25	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	<b>3</b> 93			·			603
	<210> 185						
	<211> 602						
30	<212> DNA						
	<213> Homo	sapiens					
	<400> 185				•		
	gaattctgta	agcatttcct	atgtgtacct	gcccctgggc	aaggtgggcc	tgacttgtta	60
35	gagtgttaga	gttttaccct	gttcctctag	gagggcctgg	taccaccaca	gcccagcatg	120
	gtgtggtgcc	tcagcaggag	gcatctggtt	acaatcaaca	caagctgttc	cagccaattt	180
	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
	ccaggatgtg	ccaccaatca	taaggaggca	ggggcctcct	tccgctgctc	cctgggactc	300
	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
40	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
	acttatggtg	tàctgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	gg						602

```
<210> 186
      <211> 601
      <212> DNA
      <213> Homo sapiens
5
      <400> 186
      gaattetgta agcattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
10
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420.
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
15
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
                                                                         601
      <210> 187
20
      <211> 600
      <212> DNA
      <213> Homo sapiens
      <400> 187
25
      gaattetgta agcattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
30
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccet actcetetga teaateceet gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
35
      <210> 188
      <211> 599
      <212> DNA
      <213> Homo sapiens
40
      <400> 188
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
45
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
```

```
ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
       cetetgettt aagegtaaac atggatgeec aggaeetgge etcaatette egagtetggt 420
       acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
 5
       cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaag 599
       <210> 189
       <211> 598
10
       <212> DNA
       <213> Homo sapiens
       <400> 189
      gaattetgta ageattteet atgtgtaeet geeettggge aaggtgggee tgaettgtta 60
15
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
20
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaa
25
       <210> 190
      <211> 498
       <212> DNA
      <213> Homo sapiens
3.0
      <400> 190
      taccaccaca gcccagcatg gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca 60
      caagctgttc cagccaattt aaagaaactt caggaggaat agggttttag gagggcatgg 120
      ggaccetect geaccegaag ceaggatgtg ceaceaatea taaggaggea ggggeeteet 180
      teegetgete cetgggaete tetaggtgte egtggeetea geececetet geacacetge 240
35
      atcttccttc tcatcagctt cctctgcttt aagcgtaaac atggatgccc aggacctggc 300
      ctcaatcttc cgagtctggt acttatggtg tactgacagt gtgagaccct actcctctqa 360
      tcaatcccct gggttggtga cttccctgtg caatcaatgg aagccagcga ggcagggtca 420
      catgccccgt ttagaggtyc agacttggag aaggaacgtg ggcaagtctt cccaggaaca 480
      ggtagggcag ggaggaaa
                                                                         498
40
      <210> 191
      <211> 398
      <212> DNA
      <213> Homo sapiens
```

	<400> 191						
	agggttttag	gagggcatgg.	ggaccctcct	gcacccgaag	ccaggatgtg	ccaccaatca	60
						cgtggcctca	
	gccccctct	gcacacctgc	atcttccttc	tcatcagctt'	cctctgcttt	aagcgtaaac	180
5	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	acttatggtg	tactgacagt	240
•	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	cttccctgtg	caatcaatgg	300
	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	agacttggag	aaggaacgtg	360
	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaa			398
10	<210> 192						
•	<211> 298						
	<212> DNA						
	<213> Homo	sapiens					
15	<400> 192						
	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	60
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	120
	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	180
	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	240
20	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaa	298
	<210> 193						
	<211> 198						
<b>.</b>	<212> DNA				•		
25	<213> Homo	sapiens					•
	<400> 193						
	ctcaatcttc	cgagtctggt	acttatggtg	tactgacagt	gtgagaccct	actcctctga	60
	tcaatcccct	gggttggtga	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	120
30	catgccccgt	ttagaggtgc	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	180
	ggtagggcag	ggaggaaa					198
	<210> 194						•
2.5	<211> 98						
35	<212> DNA	•					
	<213> Homo	sapiens					
	<400> 194						
	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	agacttggag	aaggaacgtg	60
40	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaa			98
	<210> 195						
	<211> 8	•					
4.5	<212> DNA						
45	-717- **						

<400> 195 8 ggaggaaa <210> 196 5 <211> 781 <212> DNA <213> Homo sapiens <400> 196 10 gggggcatct ctggtgcagc ccggttcgga gcaggaagac gcttaataaa tgctgataga 60 ctgcaggaca caggcaaagg tgctgagctg gaccctttat ttctgccctt ctcccttctg 120 gcaccccggc caggaaattg ctgcagcctt tctggaatcc cgttcatttt tcttactggt 180 ccacaaaagg ggccaaatgg aagcagcaag acctgagttc'aaattaaatc tgccaactac 240 cageteagtg aatetgggeg agtaacacaa aaettgagtg teettacetg aaaaatagag 300 15 gttagaggga tgctatgtgc cattgtgtgt gtgtgttggg ggtggggatt gggggtgatt 360 tgtgagcaat tggaggtgag ggtggagccc agtgcccagc acctatgcac tggggaccca 420 aaaaggagca tetteteatg attttatgta teagaaattg ggatggeatg teattgggae 480 agggtctttt ttcttgtatg gtggcacata aatacatgtg tcttataatt aatggtattt 540 tagatttgac gaaatatgga atattacctg ttgtgctgat cttgggcaaa ctataatatc 600 20 totgggcaaa aatgtcccca totgaaaaac agggacaacg ttcctccctc agccagccac 660 tatggggcta aaatgagacc acatctgtca agggttttgc cctcacctcc ctccctgctg 720 gatggcatcc ttggtaggca gaggtgggct tcgggcagaa caagccgtgc tgagctagga 780 781 25 <210> 197 <211> 771 <212> DNA <213> Homo sapiens 30 <400> 197 ctggtgcagc ccggttcgga gcaggaagac gcttaataaa tgctgataga ctgcaggaca 60 caggcaaagg tgctgagctg gaccctttat ttctgccctt ctcccttctg gcaccccggc 120 caggaaattg ctgcagcctt tctggaatcc cgttcatttt tcttactggt ccacaaaagg 180 ggccaaatgg aagcagcaag acctgagttc aaattaaatc tgccaactac cagctcagtg 240 35 aatctgggcg agtaacacaa aacttgagtg tccttacctg aaaaatagag gttagaggga 300 tgctatgtgc cattgtgtgt gtgtgttggg ggtggggatt gggggtgatt tgtgagcaat 360 tggaggtgag ggtggagccc agtgcccagc acctatgcac tggggaccca aaaaggagca 420 tetteteatg attttatgta teagaaattg ggatggeatg teattgggae agegtetttt 480 ttcttgtatg gtggcacata aatacatgtg tcttataatt aatggtattt tagatttgac 540 40 gaaatatgga atattacctg ttgtgctgat cttgggcaaa ctataatatc tctgggcaaa 600

771

aatgtcccca tctgaaaaac agggacaacg ttcctccctc agccagccac tatggggcta 660 aaatgagacc acatctgtca agggttttgc cctcacctcc ctccctgctg gatggcatcc 720

ttggtaggca gaggtgggct tcgggcagaa caagccgtgc tgagctagga c

```
<210> 198
       <211> 761
       <212> DNA
       <213> Homo sapiens
 5
       <400> 198
      ccggttcgga gcaggaagac gcttaataaa tgctgataga ctgcaggaca caggcaaagg 60
      tgctgagctg gaccetttat ttctgccctt ctcccttctg gcaccccggc caggaaattg 120
      ctgcagcctt tctggaatcc cgttcatttt tcttactggt ccacaaaagg ggccaaatgg 180
10
      aagcagcaag acctgagttc aaattaaatc tgccaactac cagctcagtg aatctgggcg 240
      agtaacacaa aacttgagtg teettaeetg aaaaatagag gttagaggga tgetatgtge 300
      cattgtgtgt gtgtgttggg ggtggggatt gggggtgatt tgtgagcaat tggaggtgag 360
      ggtggagccc agtgcccagc acctatgcac tggggaccca aaaaggagca tcttctcatg 420
      attttatgta tcagaaattg ggatggcatg tcattgggac agcgtctttt ttcttgtatg 480
15
      gtggcacata aatacatgtg tettataatt aatggtattt tagatttgac gaaatatgga 540
      atattacctg ttgtgctgat cttgggcaaa ctataatatc tctgggcaaa aatgtcccca 600
      totgaaaaac agggacaacg ttootcootc agccagccac tatggggota aaatgagacc 660
      acatotytoa aggyttttyc cotcacotoo otocotycty yatyycatoo ttyytagyca 720
      gaggtgggct tcgggcagaa caagccgtgc tgagctagga c
                                                                         761
20
      <210> 199
      <211> 751
      <212> DNA
      <213> Homo sapiens
25
      <400> 199
      gcaggaagac gcttaataaa tgctgataga ctgcaggaca caggcaaagg tgctgagetg 60
      gaccetttat ttetgeeett etecettetg geaccegge caggaaattg etgeageett 120
      tctggaatcc cgttcatttt tcttactggt ccacaaaagg ggccaaatgg aagcagcaag 180
30
      acctgagttc aaattaaatc tgccaactac cagctcagtg aatctgggcg agtaacacaa 240
      aacttgagtg tccttacctg aaaaatagag gttagaggga tgctatgtgc cattgtgtgt 300
      gtgtgttggg ggtggggatt gggggtgatt tgtgagcaat tggaggtgag ggtggagccc 360
      agtgcccagc acctatgcac tggggaccca aaaaggagca tcttctcatg attttatgta 420
      tcagaaattg ggatggcatg tcattgggac agcgtctttt ttcttgtatg gtggcacata 480
35
      aatacatgtg tottataatt aatggtattt tagatttgac gaaatatgga atattacctg 540
      ttgtgctgat cttgggcaaa ctataatatc tctgggcaaa aatgtcccca tctgaaaaac 600
      agggacaacg ttcctccctc agccagccac tatggggcta aaatgagacc acatctgtca 660
      agggttttgc cctcacctcc ctccctgctg gatggcatcc ttggtaggca gaggtgggct 720
      tegggeagaa caageegtge tgagetagga e
                                                                         751
40
      <210> 200
      <211> 750
      <212> DNA
      <213> Homo sapiens
45
```

```
<400> 200
      caggaagacg cttaataaat gctgatagac tgcaggacac aggcaaaggt gctgagctgg 60
      accetttatt tetgeeette teeettetgg cacceeggee aggaaattge tgeageettt 120
      ctggaatccc gttcattttt cttactggtc cacaaaaggg gccaaatgga agcagcaaga 180
 5
      cctgagttca aattaaatct gccaactacc agctcagtga atctgggcga gtaacacaaa 240
      acttgagtgt ccttacctga aaaatagagg ttagagggat gctatgtgcc attgtgtgtg 300
      tgtgttgggg gtggggattg ggggtgattt gtgagcaatt ggaggtgagg gtggagccca 360
      gtgcccagca cctatgcact ggggacccaa aaaggagcat cttctcatga ttttatgtat 420
      cagaaattgg gatggcatgt cattgggaca gcgtcttttt tcttgtatgg tggcacataa 480
10
      atacatgtgt cttataatta atggtatttt agatttgacg aaatatggaa tattacctgt 540
      tgtgctgatc ttgggcaaac tataatatct ctgggcaaaa atgtccccat ctgaaaaaca 600
      gggacaacgt tectecetea gecagecaet atggggetaa aatgagaeca catetgteaa 660
      gggttttgcc ctcacctccc tccctgctgg atggcatcct tggtaggcag aggtgggctt 720
      cgggcagaac aagccgtgct gagctaggac
                                                                         750
15
      <210> 201
      <211> 681
      <212> DNA
      <213> Homo sapiens
20
      <400> 201
      gggggcatct ctggtgcagc ccggttcgga gcaggaagac gcttaataaa tgctgataga 60
      ctgcaggaca caggcaaagg tgctgagctg gaccetttat ttctgccctt ctcccttctg 120
      gcaccccggc caggaaattg ctgcagcctt tctggaatcc cgttcatttt tcttactggt 180
25
      ccacaaaagg ggccaaatgg aagcagcaag acctgagttc aaattaaatc tgccaactac 240
      cageteagtg aatetgggeg agtaacacaa aaettgagtg teettacetg aaaaatagag 300
      gttagaggga tgctatgtgc cattgtgtgt gtgtgttggg ggtggggatt gggggtgatt 360
      tgtgagcaat tggaggtgag ggtggagccc agtgcccagc acctatgcac tggggaccca 420
      aaaaggagca tetteteatg attttatgta teagaaattg ggatggeatg teattgggae 480
30
      agcgtctttt ttcttgtatg gtggcacata aatacatgtg tcttataatt aatggtattt 540
      tagatttgac gaaatatgga atattacctg ttgtgctgat cttgggcaaa ctataatatc 600
      tetgggcaaa aatgteecca tetgaaaaac agggacaaeg tteeteecte agecagecae 660
      tatggggcta aaatgagacc a
                                                                         681
35
      <210> 202
      <211> 581
      <212> DNA
      <213> Homo sapiens
40
      <400> 202
      gggggcatct ctggtgcagc ccggttcgga gcaggaagac gcttaataaa tgctgataga 60
      ctgcaggaca caggcaaagg tgctgagctg gaccctttat ttctgccctt ctcccttctq 120
      gcaccccggc caggaaattg ctgcagcctt tctggaatcc cgttcatttt tcttactggt 180
      ccacaaaagg ggccaaatgg aagcagcaag acctgagttc aaattaaatc tgccaactac 240
45
      cageteagtg aatetgggeg agtaacacaa aacttgagtg teettacetg aaaaatagag.300
```



	•	•					
	gttagaggga	tgctatgtgc	cattgtgtgt	gtgtgttggg	ggtggggatt	gggggtgatt	360
	tgtgagcaat	tggaggtgag	ggtggagccc	agtgcccagc	acctatgcac	tggggaccca	420
	aaaaggagca	tcttctcatg	attttatgta	tcagaaattg	ggatggcatg	tcattgggac	480
						aatggtattt	
5		gaaatatgga				55	581
	<210> 203						
	<211> 481						
	<212> DNA						
10	<213> Homo	caniens			·		
10	(213) 110	Suprems					
	<400> 203						
							60
						tgctgataga	
16						ctcccttctg	
15						tcttactggt	
						tgccaactac	
						aaaaatagag	
					_	gggggtgatt	
						tggggaccca	
20	aaaaggagca	tcttctcatg	attttatgta	tcagaaattg	ggatggcatg	tcattgggac	480
	a						481
		• •			•		
	<210> 204						
	<211> 381						
25	<212> DNA		•				
	<213> Homo	sapiens			•	•	
							•
	<400> 204						
	gggggcatct	ctggtgcagc	ccggttcgga	gcaggaagac	gcttaataaa	tgctgataga	60
30						ctcccttctg	
						tcttactggt	
						tgccaactac	
				•		aaaaatagag	
						gggggtgatt	
35		tggaggtgag		3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	33 - 3333	33333-3	381
		- 5555.55		•			
	<210> 205	•	•		•		
	<211> 281						
	<212> DNA						
40	<213> Homo	canienc					
<del>-10</del>	-2137 MUNIO	adicus			•		
	-400- 205						
	<400> 205			~~~		h	60
٠.						tgctgataga	
45						ctcccttctg	
73	geaeeeegge		ecgeageett	cctggaatcc	egtteattt	tcttactggt	180



	•	ggccaaatgg aatctgggcg	•			tgccaactac	240 281
5	<210> 206 <211> 181 <212> DNA <213> Homo	sapiens					
10	ctgcaggaca	ctggtgcagc caggcaaagg caggaaattg	tgctgagctg	gaccctttat	ttctgccctt	ctcccttctg	120
	<210> 207 <211> 81 <212> DNA <213> Homo	sapiens					
20		ctggtgcagc caggcaaagg		gcaggaagac	gcttaataaa	tgctgataga	60 81
25	<210> 208 <211> 38 <212> DNA <213> Homo	sapiens			·		
30	<400> 208	ctggtgcagc	ccggttcgga	gcaggaag			38